

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher: H. S. SmithNA Sequence (#) 20

STN

Searcher Phone #: _____

AA Sequence (#) 10

Dialog

Searcher Location: _____

Structure (#) _____

Questel/Orbit

Date Searcher Picked Up: _____

Bibliographic

Dr.Link

Date Completed: 5/24/2000

Litigation

Lexis/Nexis

Searcher Prep & Review Time: 25

Fulltext

Sequence Systems

Clerical Prep Time: 11

Patent Family

WWW/Internet

Online Time: 22

Other

Other (specify)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:02:22 ; Search time 6115.43 Seconds
(without alignments)
-222.382 Million cell updates/sec

Title: US-09-215-435-71
Perfect score: 1398
Sequence: 1 gatgctgagggcccgagc.....acattgaaaaaaaaaaaaa 1398

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sts:*
- 14: gb_sy:*
- 15: gb_un:*
- 16: gb_v1:*
- 17: em_fun:*
- 18: em_hum1:*
- 19: em_hum2:*
- 20: em_in:*
- 21: em_om:*
- 22: em_or:*
- 23: em_ov:*
- 24: em_pat:*
- 25: em_ph:*
- 26: em_pl:*
- 27: em_ro:*
- 28: em_sts:*
- 29: em_sy:*
- 30: em_un:*
- 31: em_v1:*
- 32: gb_htg1:*
- 33: gb_htg2:*
- 34: gb_in1:*
- 35: gb_in2:*
- 36: em_ba1:*
- 37: em_ba2:*
- 38: em_hum3:*
- 39: em_hum4:*
- 40: gb_pr4:*
- 41: gb_htg3:*
- 42: gb_htg4:*
- 43: gb_htg5:*
- 44: gb_htg6:*

- 45: gb_htg7:*
- 46: em_htg1:*
- 47: em_htg2:*
- 48: em_htg3:*
- 49: em_hum5:*
- 50: gb_pl3:*
- 51: gb_pr5:*
- 52: gb_htg8:*
- 53: gb_htg9:*
- 54: gb_htg10:*
- 55: gb_htg11:*
- 56: gb_htg12:*
- 57: gb_htg13:*
- 58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	359.8	25.7	461	13	G13307	G13307 human STS W
C 2	46.8	3.3	7218	5	I66494	I66494 Sequence 14
C 3	46.4	3.3	1639	11	HSU82987	U82987 Human Bcl-2
C 4	45	3.2	1471	4	XLAPEG	X51394 Xenopus mRN
C 5	45	3.2	43501	11	AC005783	AC005783 Homo sapi
C 6	44.8	3.2	1620	1	AB012768	AB012768 Pseudomon
C 7	44.8	3.2	4261	1	AB016241	AB016241 Streptomy
C 8	44.6	3.2	4068	1	SCP6PP	X92520 S.coelicolo
C 9	44.6	3.2	32746	42	AC013224	AC013224 Drosophila
C 10	44.6	3.2	38640	2	SCL2	AL137778 Streptomy
C 11	44	3.1	12363	2	AE002088	AE002088 Deinococc
C 12	43.6	3.1	1957	5	I62451	I62451 Sequence 3
C 13	43.6	3.1	2657	4	GG073678	U73678 Gallus gall
C 14	43.6	3.1	102713	44	AC019126	AC019126 Homo sapi
C 15	43.6	3.1	128779	10	HS29K1	Z98745 Human DNA S
C 16	43.6	3.1	156601	10	HS21266	AL005172 Homo sapi
C 17	43.6	3.1	187543	40	AC005678	AC005678 Homo sapi
C 18	43.2	3.1	11905	2	AP033059	U33059 Actinosyne
C 19	43	3.1	4278	34	DMVP3G	X04754 Drosophila
C 20	43	3.1	4280	34	DR0YP3	M15898 D.melanog
C 21	43	3.1	92969	55	AC010844	AC010844 Drosophila
C 22	43	3.1	168266	43	AC011069	AC011069 Drosophila
C 23	42.8	3.1	5400	1	AVIHUP2	L25315 Azotobacter
C 24	42.8	3.1	78172	7	AB010070	AB010070 Arabidops
C 25	42.8	3.1	121882	44	AC019147	AC019147 Homo sapi
C 26	42.8	3.1	190943	11	CNS01DV1	AL133445 Human chr
C 27	42.2	3.0	6643	40	AF163302	AF163302 Homo sapi
C 28	42.2	3.0	171480	44	AC021165	AC021165 Homo sapi
C 29	42.2	3.0	192326	55	AC022519	AC022519 Homo sapi
C 30	42	3.0	4567	4	CHKCSEAX	L12024 Gallus gall
C 31	42	3.0	69350	1	MTV004	AL009198 Mycobacte
C 32	42	3.0	137101	43	AC015663	AC015663 Homo sapi
C 33	42	3.0	147853	45	AC019011	AC019011 Homo sapi
C 34	42	3.0	170301	55	AC022457	AC022457 Oryza sat
C 35	42	3.0	183871	41	AC007365	AC007365 Homo sapi
C 36	42	3.0	241625	44	AC018924	AC018924 Homo sapi
C 37	41.8	3.0	2378	11	HSU64454	U64454 Human 3' of
C 38	41.4	3.0	756	5	AR014110	AR014110 Sequence
C 39	41.4	3.0	2548	4	GGHSP905	X15028 Chicken hsp
C 40	41.4	3.0	15742	2	AF013216	AF013216 Myxococcu
C 41	41.4	3.0	152925	45	AC011199	AC011199 Homo sapi
C 42	41.2	2.9	15168	1	PSNARXL	Y15252 Pseudomonas
C 43	41.2	2.9	25359	34	CEF02D10	Z67990 Caenorhabdi
C 44	41.2	2.9	87184	41	AC009185	AC009185 Homo sapi
C 45	41.2	2.9	182586	41	AC009719	AC009719 Homo sapi

ALIGNMENTS

Db	399	TTCCCTAAGAGGACTGCAACCCAGCGGCTTGAGGGGACCCACCTTCCAGCAGGACCCCG	340
QY	1051	agctccacagtg---ccaagaagggggagggaagggggagcaggcagcctctggcca	1107
Db	339	AAGTCCCCANAGTGGCCCCAAGAAGGGGAGGAGGAAGGGGCGACAGGCCACCTCTGGCCA	280
QY	1108	ctgcagaccccggaaggtcaaggtcacatccatcctctgaaaccagagggagcctcagc	1167
Db	279	CTGCAGACCCGGAGGTCAAGGCTGACATCCATCCCTTGGAAACAGAGGGGACCTCAGC	220
QY	1168	ctcttagcaggaggtctctctgttgcaactcacccctctcttattgtctgcctgcgcac	1227
Db	219	CTCTTAGCAGAGGCTCTCTTGTTCGACTCACCTTCTTAATTGCTTGCCTCGCATC	160
QY	1228	tgggggtctgaatttttggagacaggaacaatctctgaagtgcaaaa-caggccctacgct	1286
Db	159	TGGGGTCTGAAATTTTGGGACGAGCAATATCTGAAGGTGCAAAACAGGCCCTTACGGCT	100
QY	1287	gttccctgcacaaactcctatggttttaattgtaccoccatctccacatctttaaaagctca	1346
Db	99	GTTCCCTGCACAACTCTCATGCTTTTAATTGTATGCCCATCTTCACATCTTTAAAGCTCA	40
QY	1347	tgtgaaaatgctgaatttttaataaataaactgataatttg	1385
Db	39	TGTGAAAAATGCTGCATTTTAAATAAAATGATACATTTG	1
RESULT	2		
LOCUS	166494/c		
DEFINITION	166494	7218 bp DNA	PAT
ACCESSION	166494	Sequence 14 from patent US 5670367.	
VERSION	166494.1	GI:2724471	
KEYWORDS		Unknown.	
SOURCE		Unknown.	
ORGANISM		Unclassified.	
REFERENCE		1 (bases 1 to 7218)	
AUTHORS		Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.	
TITLE		Recombinant fowlpox virus	
JOURNAL		Patent: US 5670367-A 14 23-SEP-1997;	
FEATURES		Location/Qualifiers	
source		1..7218	
BASE COUNT	1944 a	1491 c	1486 g
ORIGIN		1929 t	368 others
Query Match		3.3%;	Score 46.8; DB 5; Length 7218;
Best Local Similarity		5.1%;	Pred. No. 0.95;
Matches	21;	Conservative 215;	Mismatches 172; Indels 0; Gaps
QY	715	gttgggggcagagagctcagcgctcagagagcgggagagagagcttctgcctttcagc	774
Db	1438	GTACRR	1379
QY	775	ctggctgcgctgctatggcatgcaggacatgagctcctcaggaecggcgtatgcgcgtac	834
Db	1378	RR	1319
QY	835	catctggttcagggggacgtcagcgttggcacccaaagggcgcaagtcgcgcaaaa	894
Db	1318	RR	1259
QY	895	gaatccaaggccacacagtgcctgaggacagatggaggagcgtttgcctccgag	954
Db	1258	RR	1199
QY	955	caagcccttcaggacacagaggcaagagagacttctctagagagactgcaaccca	1014
Db	1198	RR	1139
QY	1015	gcggcctgaggggaccagcctccagcaggaccagagctcccaagctcccaagtcaccaagaagg	1074

[illegible]


```

/rpt_family="MIR"
complement(10315..10616)
/rpt_family="AluSx"
11194..11614
/rpt_family="L1PA12"
complement(11616..11908)
/rpt_family="AluSg"
12237..12484
/rpt_family="L1MB7"
14005..14181
/rpt_family="MLT1C"
14182..14483
/rpt_family="AluSx"
14499..14718
/rpt_family="MLT1C"
14960..15260
/rpt_family="AluJo"
15717..15900
/rpt_family="MER20"
complement(16772..17028)
/notes="predicted exon, program: groll2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000"
complement(17253..17381)
/rpt_family="FLAM_C"
17447..17670
/rpt_family="LINE2"
complement(17718..18013)
/rpt_family="AluSg"
18187..18271
/rpt_family="LINE2"
19047..19347
/rpt_family="AluY"
complement(19410..19515)
/rpt_family="MIR"
complement(20778..20911)
/rpt_family="FLAM_C"
21752..22037
/rpt_family="AluSg"
22038..22061
/rpt_family="(TAA)n"
22167..22373
/rpt_family="MER20"
22574..22678
/notes="predicted exon, program: groll2exons_human_1.3,
frame: 2, quality: excellent, score: 82.000"
23016..23199
/rpt_family="LINE2"
23287..23494
/rpt_family="MER20"
23740..23928
/rpt_family="MER20"
23946..24060
/rpt_family="LINE2"
24191..24485
/rpt_family="AluSx"
24575..24797
/rpt_family="MIR"
complement(25744..25782)
/rpt_family="tRNA-Gly-GGA"
complement(26326..26386)
/rpt_family="GC-rich"
complement(26579..26852)
/notes="BLASTN similarity to X64281 (1005..1279); match:
0.99, score: 2.6e-95; database searched: nt; H.sapiens
variant gene for tRNA-Val (anticodon:CAC) and tRNA-Gly
(anticodon:UCC)"
complement(26589..26625)
/rpt_family="GC-rich"
26761..26832
/rpt_family="tRNA-Gly-GGA"
complement(26860..27051)
/rpt_family="AluSx"
27052..27087

```

```

repeat_region      /rpt_family="POLY G"
                    complement(27088..27187)
repeat_region      /rpt_family="AluSx"
                    27206..27418
                    /note="BLASTN similarity to Z59776 (1..214); match: 0.98,
Query Match      3.2%; Score 45; DB 11; Length 43501;
Best Local Similarity 49.4%; Pred. No. 2.5;
Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 99 gccgaacccctgaggtgccctttgagagcagtgctacacgcatctcagcttcagccgcg 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26515 GCCTGAGCGCGCGGTCCCTGTGTCTCCCGCGGTGTCTCCCGCTCCCGCACTTC 26456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 gcaagagctgcgcctatactagacctgtgcctggggccagcccaaacagagccac 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26455 CGGGTGTGAGTCCCGGCATGGAGCCGCGACCGGTGAGCGCGCTGGACCGGACGAC 26396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 tggccctggtcttcgcttcgcatgtccggtctcttttcagctggtgccccgcgaggagc 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26395 CTGCCAGTCCGCGCGCGCCACGTCCCGGTGAGCAGCGCGCGCTCGTCTCCGCGC 26336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 tgcacgcatgcccacacctgcgctttacacgcccccccgctggccccgcgcgcc 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26335 CCCGCGCGGACCCCTGAACCGCCATTGACGGCGCGCGCTCGGCCCTGACCCCTGCC 26279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
AB012768          1620 bp      DNA      BCT      09-OCT-1998
LOCUS             Pseudomonas aeruginosa gene for Cher, complete cds.
DEFINITION        AB012768
ACCESSION         AB012768
VERSION           AB012768.1 GI:3721577
KEYWORDS          Cher.
SOURCE            Pseudomonas aeruginosa (strain:PAO1) DNA.
ORGANISM          Pseudomonas aeruginosa
                  Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
                  Pseudomonas.
REFERENCE          1 (bases 1 to 1620)
AUTHORS           Kato,J., Ohtake,H. and Akio,K.
TITLES            Direct Submission
JOURNALS           Submitted (06-APR-1998) to the DDBJ/EMBL/GenBank databases. Junichi
                  Kato, Hiroshima University, Department of Fermentation Technology;
                  1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan
                  (E-mail:jun@ipc.hiroshima-u.ac.jp, Tel:81-824-24-7757,
                  Fax:81-824-22-3758)
REFERENCE          2 (sites)
AUTHORS           Kato,J., Nakamura,T., Kuroda,A. and Ohtake,H.
TITLES            Cloning, sequence and characterization of chemotaxis genes in
                  Pseudomonas aeruginosa
JOURNALS           Unpublished (1998)
FEATURES           Location/Qualifiers
                    1..1620
                    /organism="Pseudomonas aeruginosa"
                    /strain="PAO1"
                    /db_xref="taxon:287"
                    235..1059
                    /gene="cher"
                    235..1059
                    /gene="cher"
                    /function="methyltransferase"
                    /note="cher is involved in methylation of transducer
                    protein in chemotaxis signal transduction. cher gene
                    locates at more than 1.5 Mb distance from chemotaxis gene
                    cluster."
                    /codon_start=1
                    /transl_table=11
                    /product="Cher"
                    /protein_id="BAA33556.1"
                    /db_xref="GI:3721578"
                    /translation="MSAANADFLFRVLEKTCGIVLGSNNQYLVSSLNKLMBQOGI
                    KSLGELVRIQTQRGLREVMVDANTNETLWFRDTPYFVLKORVLPFLIKANGGOR
                    LRIWSAACSSSQEPYSLSAIDFEKTNLGQLKAGVQIVATDLSSLTAAKACEYDT

```

LAMRGSLSPERLQRYFDAKPGRWAKPAIRSRVFRALNLLDSYASLGKFDMMVFCRN
VLIFYSAEVRDILLRHGTLKPGGYLFLGSEALNLPDHYQWQCSPGIIYRAK"

BASE COUNT
ORIGIN

288 a 509 c 536 g 287 t

Query Match 3.2%; Score 44.8; DB 1; Length 1620;
Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 118; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy 483 tctgcagccctctgacagaggttcttcaatgagcattggaactatctgcccag 542
Db 263 TGTTCAGGCTTCTCTCGAAGACCTGCGGCATTTGCTGGGAGCAACAGCAGTACC 322

Qy 543 agatcctgtacggctgaagatccccctttgagaagggccgctcggtctctgagcccc 602
Db 323 TGGTCTCGAGCGGCTGACACAGCTGATGGAGCAGGCGCATCAAGAGCTGGGAGAGC 382

Qy 603 tgcagcgcacagccgagccgagctgagctgacccagagatagccagagatgaagaccagctgc 662
Db 383 TGGTGCAGCGCATCCAGACCCAGCGCGGCGCTGCGGAGATGGTGGTGGACGCGATGA 442

Qy 663 agaaccagaccctgagctatgtcaactcagtgcccaaggaagtggtccagttggggg 722
Db 443 CCACCAAGAGACCCCTGTGGTTTCGCGATACCTACCCGTTTCGAGGCTCTCAAGCAGCGGG 502

RESULT 7

AB016241 4261 bp DNA BCT 06-FEB-1999
LOCUS Streptomyces castaneoglobisporus plasmid pSY10 DNA for replication
DEFINITION protein, complete cds.
ACCESSION AB016241
VERSION AB016241.1 GI:3347665
KEYWORDS rep; replication protein; spd B2; transfer within recipient cell
BI.

SOURCE Streptomyces castaneoglobisporus plasmid: pSY10 DNA.
ORGANISM Streptomyces castaneoglobisporus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 4261)
AUTHORS Suzuki, K., Sugiyama, M. and Ikeda, K.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1998) to the DDBJ/EMBL/GenBank databases. Koji
Suzuki, Ibaraki National College of Technology, Dept. of Chemistry
and Material Engineering, Nakane 866, Hitachinaka, Ibaraki
312-8508, Japan (E-mail: k.suzuki@cc.ibaraki-ct.ac.jp,
Tel: 81-29-271-2989, Fax: 81-29-271-2992)

REFERENCE
AUTHORS

Ikeda, K., Suzuki, K., Yoshioka, H., Miyamoto, K., Masujima, T. and
Sugiyama, M.

TITLE Construction of a new cloning vector utilizing a cryptic plasmid
and the highly expressed melanin-synthesizing gene operon from
Streptomyces castaneoglobisporus

JOURNAL FEMS Microbiol. Lett. 168, 195-199 (1998)

FEATURES

Location/Qualifiers
1..4261

/organism="Streptomyces castaneoglobisporus"

/plasmid="pSY10"

/db_xref="taxon:79261"

271..552

/gene="spd B1"

271..552

/gene="spd B1"

/function="involved in transfer of the plasmid within a
recipient cell"

/codon_start=1

/transl_table=11

/protein_id="BAA31970.1"

/db_xref="GI:3347666"

/translation="MQRSTAAAPAVAMTAVSMVLTVAVVMWLGSMPPVAVVLGLG
IDGWLATLAYDRHLGVRGRDRRPRGRAGRLARPVAGRRLAPARG"

695..1114

gene

CDS

/gene="spd B2"
695..1114
/gene="spd B2"
/function="involved in transfer of the plasmid within a
recipient cell"
/codon_start=1
/transl_table=11
/protein_id="BAA31971.1"
/db_xref="GI:3347667"

CDS

/note="orf 68"
/codon_start=1
/transl_table=11
/protein_id="BAA31972.1"
/db_xref="GI:3347668"
/translation="MTALAYLTAVAVGLVLAVALRPRTAPPAPVPAPAPALPA
TAAETRAILVAAQLAARLDTAA"

gene

2064..3581
/gene="rep"
2064..3581
/gene="rep"

CDS

/note="a protein essential for replication"
/codon_start=1
/transl_table=11
/product="replication protein"
/protein_id="BAA31973.1"
/db_xref="GI:3347669"

/translation="MLDPTGVYIGRTARGVVTGLVACBRIFWCSCSAAIRIGRT
RELTKAALRLAAGTTLAVVLTARHRTRLVLSDALMGAPKLDAGSPVLDHSGK
PRRDGQRMLTDPAFGRPEALNTWRKDSVGTSVRPVPGVIRHIGVIMVRAS
EYTRRENGWHPHMLLAFLGTTLEGSPAKVYVTHGHTPGQEALEDWEDWLRALTRT
LRADAFEPETKCGIOGCKGRGHVYVRIISADDAALIEYLVKVDGKSGASV
QEDLMAAGAAETVRADNKRGRDMSVFPOMLYRLMDLAEGDDEQAEGYCTALO
LRDWAIEYQAMSGRRALTEWTLGLRRHVLGDDSEERDSEVYBOAAREPLTCGVYM
TTTAYKVVGKLAPEVEEIVAAEVYDSVADLVGLGLAGHVRPLTAQELADQEHM
AARMLAQREDREKQARIAQLADQANATERARRRFGALLVQLLAARHTGTEPSSH"

BASE COUNT 605 a 1538 c 1559 g 559 t
ORIGIN

Query Match 3.2%; Score 44.8; DB 1; Length 4261;
Best Local Similarity 44.1%; Pred. No. 2.7; 237; Indels 0; Gaps 0;

Matches 187; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 396 gccgcgggcccctgtgtcttgcaggaggtaccagcagttcaggaggagaatgtgctacgaaacc 455

Db 369 GCCCTGGCCCGTCGCCCTGGCTCGGCTCGGCATCGACGGCGCATGGCTGCCACCCCT 428

QY 456 tagcgataagcccttgaccggcccatctgcaggccctctgcagcagaggttcttca 515

Db 429 CGCCTACACCGGCATCGGCTGGCGGTTTCGGCGCGGTCGCCGACCGCGCTCCTGGTCTCC 488

QY 516 atggcatigggcaactatctgcgggcagagatcctctacggctgagatccccctttg 575

Db 489 AGCCCTGGCGGAGGACCTCGCCGCCCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 548

QY 576 agaagggcccctcggtctcttgcaggccctgcagcagcacagccgagccgagctgacccc 635

Db 549 CTAGGCTCTCTGGCTCGTCCACGGCTCTGGGAGCGCACCGCGCTCACCGACACGCCCC 608

QY 636 tgagccagaagaataaggaaccaagctgcagaatccagactgagagctgagagctatgtcaactag 695

Db 609 TGGACCATCGCCACCATCCAGAGAGAGCCCGCCGACCGCGCCATCCCGCGCGCG 668

QY 696 tgcacaaggaagtgttccagttgggggcagagagctcagggctcagagagcgaggaggg 755

Db 669 AGCTGCGGGCGGAGCGGCGGAGGAGTGACCGCGCTGACGCGCGCTGACGCGCGCGGG 728

QY 756 actttgtgcctttcagagcctggtgcgctgctatggaatccagggagatgagctccctgc 815

Db 729 CCCGCTGCCAGCCCGAGTCCCGGGCCGCCAGCGCTGGCCGGCGCTGGTGGGAGC 788
QY 816 agga 819
|||
Db 789 TGGa 792
|||
RESULT 8
SCPPGPP 4068 bp DNA BCT 29-AUG-1996
LOCUS S.coelicolor orf1, orf2, and orf3 genes.
DEFINITION X92520
ACCESSION X92520
VERSION X92520.1 GI:1314143
KEYWORDS (p)ppGpp synthetase; adenine phosphoribosyltransferase; matrix metalloproteinase; MMP-18 gene.
SOURCE Streptomyces coelicolor.
ORGANISM Streptomyces coelicolor
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 4068)
AUTHORS Martinez-Costa O.H., Arias, P., Romero, N.M., Parro, V., Mellado, R.P. and Malpartida, F.
TITLE A relA/spoT homologous gene from Streptomyces coelicolor A3(2) controls antibiotic biosynthetic genes
J. Biol. Chem. 271 (18), 10627-10634 (1996)
REFERENCE 96209989
2 (bases 1 to 4068)
AUTHORS Malpartida, F.
TITLE Direct Submission
Submitted (24-OCT-1995) F. Malpartida, Centro Nacional de Biotecnologia, Campus de la Universidad Autonoma, 28049-Canto Blanco, Madrid, SPAIN
JOURNAL Location/Qualifiers
FEATURES
source 1. .4068
/organism="Streptomyces coelicolor"
/strain="A3 (2) J802"
/db_xref="taxon:1902"
gene 1. .127
/gene="orf3"
CDS <1. .127
/gene="orf3"
/codon_start=2
/transl_table=11
/product="adenine phosphoribosyltransferase"
protein_id="CAA63298.1"
/db_xref="GI:1314144"
RBS /db_xref="SWISS-PROT:P52561"
/translation="IRRAGAEVAGLAVLMELGFLGGRARLEPALAGAPLEALLTV"
300. .3005
/gene="orf1"
gene 300. .3005
/gene="orf1"
CDS 311. .2854
/gene="orf1"
/codon_start=1
/transl_table=11
/product="(p)ppGpp synthetase"
protein_id="CAA63297.1"
/db_xref="GI:1515319"
/db_xref="SPTREMBL:P72401"
/translation="MPDEAOPLTAAKPESASAAKPAAPSAQAKNDTHGP1QHAPAA
PVDPKAEQPPKPLPAERPNQAPVVRAPAGAPARSGSSNVRARLARLVQRANPN
PVLPLRLVRGNDPKLETSLROIERYAQVAERWHRGQRKSGDPYTHPLVATLIL
AELGMDPATLMAGLLHDTVEDTEYGLDRDGLVTLTLVGVTKLDRKVFGEAAQA
ETVRKMYAMAKDRVLVLIKLAIDLHNNRTMYLRKREKKEKARETLEIYAPLAHRLG
MNTIKWLEDLAFALPKMYDEIVRLVAERAPKRDVLAWTVDEQODLRAARIKAT
VTGRPKYISVYQKRVGRDFAEIVDLVGLRVLDVTRDCYAALGVTHARWNPVGR
FKDYIAMPKKNMYQSLHTTVIGPGGKPEVLQIRTFDMHRRAYEGIAHWWKQKQAVG
ASKVTDAPKSGSKSDHNLDMWLKQLDQKRETDPPGEFLSLRFDLSRNEVVF
TRGDVTLAPAGTVPDFAVHTEVGHRTIGARVNGRLVPLESTLNDGLVFEVTSK
AAGAGPSRDLGFKYSPRANKIRAWFSKERRDEAI EQGDAIVRAMRKQLPLQRI
TGDSLVTLAHEMRYSDISALYAAIGEGHVSAPNIVQKLVQALGEGEAATEIDESVPP
SRGRGRRRANADPGVVVKGVEDVWVKLARCCCTPVPDPIIGFVTRGSGVSVHRSDCV

mRNA NVDSLREPERILEVEWAPQTSSVFLVAIQVEALDRSLLSLDVTRVLSQHVNLISAA
VQTSRDRVATSRFTFEMGDPKHLGHVLKAVRGVEGVYDVYRVTSARRPS"
2933. .>4068
/gene="orf2"
gene 2933. .4068
/gene="orf2"
BASE COUNT 607 a 1530 c 1373 g 558 t
ORIGIN
Query Match 3.2%; Score 44.6; DB 1; Length 4068;
Best Local Similarity 48.6%; Pred. No. 3;
Matches 122; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 167 ctggcctgatactagccctctgctggggccagcccaacagagcactgcccctg 226
Db 3737 CGCCACTAGTTCGAGCGCGCAGCTCTTCGGCCTCGGGGACGAGGCTCTTGCGGC 3796
QY 227 gtctcccgcttcgcatgtccgctcttttcagctggtgccccgcgagagctgccacgc 286
Db 3797 GCGCGCTCGTCGGACTGCTTGGCCGCTGGGCTTCCTCGCGCGGGTCTCCAC 3856
QY 287 catgccccctgcgcttttaacagcccccgcctggccccggctgcctatgttctg 346
Db 3857 GAGCGGACGAGTGGTCCAGCGCGCGCCGACGAGGCTCCAGGTGCGCCGCGTGGTG 3916
QY 347 gacatccgcttcgctggcgcgtggacacctgggggaaagtggcagccggcggggcc 406
Db 3917 CGCGTCCACCTGCTCGCGCAGGTGTCGACGCGGCTCTTGGCGCTCCTTGGCCGACAGGTC 3976
QY 407 tgtgtcttgca 417
Db 3977 GGTGCTCTTCA 3987
RESULT 9
AC013224/c 32746 bp DNA HTG 03-NOV-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC013224
VERSION AC013224.1 GI:6223108
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 32746)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10214289 by the submitter.
For further information on this sequence you may e-mail to
fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source Location/Qualifiers
1. .32746
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 9569 a 6832 c 6980 g 9365 t
ORIGIN
Query Match 3.2%; Score 44.6; DB 42; Length 32746;
Best Local Similarity 52.4%; Pred. No. 3;
Matches 98; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 575 gagaagggccgctcgctcgagccctgagccctgcagcagcacagccgagccggagctgacc 634

/db_xref="GI:6822209"
/translation="MPDEAQLTAAPKESASAAKPSAPQAKNDTHGPIQHAPAA
PVDKPAEQPRPKPLPAERPNAPVVRAPAGAPARGSSNRVRARLARLGVRPNYN
PVLPLELLRVGRNDPKLSTLRQERAYQVAERHWRGOKRSGDPYITHPLAVTIL
AELGMDPATLMAGLLHDVTEDDEYGLEDRDFGVVTLVGVTKLDRVKFGERAQA
ETLRKVMAMADPRVLVTKLADRLHNMRTMYLRKEQEKARETLEYAHLRLG
MNTIKWELDEAFALYPKMYDEIVRLVAERPKDEYLAUVYDEVQDLRAARKAT
VTGRPKHYYSVTKMIVRDEFAEIIYDLVIRVLDVRDCYAAAGTVHARNVPVGR
FKDYTAMPFNQSKLHTTVIGPGKPVLOJRTFDMHRAEYVGAHMKYKOEAVG
ASKVRDAPKSSQKDLNDMLRQLLDQKTEDEPGEFLESIRFDLSRNEVFV
TPKGDVIALPAGATPVDFAYAVHTEVGHRTIGARVNGRLVPLESLDNGDLVEVTSK
AAGSPSRDLGVFSYARNKIRAWFSKREDEAEQGDALIVRAARKNOLPIQRIL
TGDSLVLAHMYRISDISALYAAIGEGHVSAPNIYQKLQALGGEAAATEEDIESPP
SRGRKRRANADPGVVVGVGDEYVWKLARCTPVPDPIIGFVTRGSGVSRHSDCV
NVDSLRSRPERILEVWAPTOSSVFLVAIQEALDRSLSDVTRVLSQHNILSAA
VQTSRDRVATSRFTMEGPKHLGHVILKAVRGEVYDVYRVSARRPS"
/genes="relA"
/note="Pfam match to entry PF01842 ACT, ACT domain, score
63.10, E-value 6e-15"
complement(32378..2401)
/genes="relA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(3273..7460)
/note="previously sequenced region SW:SCSECAPT EMBL:X85969
S.coelicolor secD, secF & apt genes"
complement(3778..3781)
complement(3904..4568)
/note="previously sequenced region SW:SCAPTRELA
EMBL:X87267 S.coelicolor apt & relA genes"
complement(3956..4453)
/genes="apt"
complement(3956..4453)
/genes="apt"
/note="SCl2.04c, apt, adenine phosphoribosyltransferase,
len: 182 aa; identical to previously sequenced
SW:APT_STRCO (EMBL:X87267) Streptomyces coelicolor adenine
phosphoribosyltransferase (EC 2.4.2.7) Apt, 182 aa and
highly similar to SW:APT_ECOLI (EMBL:M14040) Escherichia
coli adenine phosphoribosyltransferase (EC 2.4.2.7) Apt,
183 aa; fasta scores: opt: 600 z-score: 682.8 (E):
1.2e-30; 54.8% identity in 168 aa overlap. Contains Pfam
match to entry PF00156 Priboyltran, Phosphoribosyl
transferase domain and match to prosite entry PS00103
Purine/pyrimidine phosphoribosyl transferases signature"
/codon_start=1
/transl_table=11
/product="adenine phosphoribosyltransferase"
/protein_id="CAB70916.1"
/db_xref="GI:6822210"
/translation="MADYEPGVVFKDITPLLADPCAFPAALTDALAEAGRATKVV
GLEARFILGAPVALRGLGIPVRKAGLPGATLSQAYDLEYGSAEIEVHADLITAG
DRVLVDDVLTAGGTAEASLELIRRAGAEVAGLVLMEILGGLGRARLEPALAGAPLE
ALITV"
/genes="apt"
complement(3980..4429)
/note="Pfam match to entry PF00156 Priboyltran,
3.5e-41"
complement(4103..4141)
/genes="apt"
/note="PS00103 Purine/pyrimidine phosphoribosyl
transferases signature"
complement(4468..4503)
complement(4501..5622)
/genes="secF"
complement(4501..5622)
/genes="secF"
/note="SCl2.05c, secF, protein-export membrane protein,
len: 373 aa; identical to previously sequenced
SW:SECF_STRCO (EMBL:X85969) Streptomyces coelicolor
protein-export membrane protein SecF, 373 aa and similar

Query Match 3.2%; Score 44.6; DB 2; Length 38640;
Best Local Similarity 48.6%; Pred. No. 3;
Matches 122; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 167 ctgcgcctgtactagccctctctctggggccagcccaacaggagccactggccctg 226
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 CCGCACCTGGTCGAGCGCGCCACGCTCTCGGCCTCGCGCAGCCAGGCGCTCTTGGGCGC 288
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 227 gtcttcgccttcgcgcctctcttcacagccgcctctgtgccccgcgagagctccacgc 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 GCGGCCCTCGTCGGACTGCTTGGCCCCCTGGGCTCTCGGCTCTCGCGCGGGTCTCCAC 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 287 catgccacacctggcctttacacagccgcctctggcccccgcctgcgcctatgttctg 346
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227 GAGCGCGACGAGCTGGTCCAGCGCGGCCCGCAGGCGCTCCAGGTCTCGCGACCGGTGGTG 168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 347 gacatccgcggttcgcgcctctgggacctgggggaagtggcagccggcgccggccccc 406
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 167 CGGTCCACCTGCTCGCGCAGGTGTGCGAGCGGCTGTGGCGGTCTTGGCGCGACAGTC 108
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 407 tgtgtcttgca 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 GGTGGTCTTCA 97
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 11
AE002088
LOCUS
DEFINITION Deinococcus radiodurans RI section 225 of 229 of the complete
chromosome 1.
ACCESSION AE002088 AE000513
VERSION AE002088.1 GI:6460412
KEYWORDS
SOURCE Deinococcus radiodurans.
ORGANISM Deinococcus radiodurans.
REFERENCE 1 (bases 1 to 12363)
Bacteria; Thermus/Deinococcus group; Deinococcus.
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE Genome sequence of the radioresistant bacterium Deinococcus
radiodurans RI
JOURNAL Science 286 (5444), 1571-1577 (1999)
MEDLINE 20036896
REFERENCE 2 (bases 1 to 12363)
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 12363
/organism="Deinococcus radiodurans"
/strain="RI"
/db_xref="taxon:1299"
/chromosome="1"
complement(94..798)
/genes="DR2585"
complement(94..798)
/genes="DR2585"
/note="similar to GB:AE000657 percent identity: 56.65;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
gene
CDS

IGDVPDKAARLTSLGDTSRMHFGLLPRANRIFAVNELADLSPKQVVALFNILQEG
DVOIKGVPIRLEDVDMVFSANPEDYTARGKIVTPLKDRIGSCIRTHPYTDVGLGMDI
TAQEAKEGVVPPFIASLIEDIAFOAREDEGRIDKLSGVSORLPISLAMELANAER
RSLSVGDAPVRVDYVAGLPALTIGKMELEYEGELGADVAKDLIRKAAGAIYAKSY
GSADTRELEKWFQGNVFRFPQGDASALKATGEVPGJSDFAAQVAGSSDDAVRVSA
AEFILDGLYGRKKLSRAEELYAAPEPELKRGRWN"
complement(8608..9618)
/gene="DR2595"

Query Match 3.1%; Score 44; DB 2; Length 12363;
Best Local Similarity 50.5%; Pred. No. 4.2;
Matches 107; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
Qy 222 ccttggtctccgttcgcatgccgctcttttcagctgggtgcgccgagagctgc 281
Db 1168 CGCTGGGTACCGCTTCGTGGTGATGTCACCTTTTTCCTGTGCTGCTACGGGAATGC 1227
Qy 282 cagccatgccacctgcgtcttttacacgccccgcctggccccgcctgcctatgtt 341
Db 1228 ATATCGAGTGCACGGCGGAGCACGTGCGCGCGCCGACCCGCTGCTGCTGGCCG 1287
Qy 342 tctggaacatccgcggttcgcgcctggagaccttgggggaaagtggcagccggccgcg 401
Db 1288 CCAACACCACTCGCGCTGCACCCCTTCTGTGATGCGCGCGCTGCGCGCGCGCTT 1347
Qy 402 ggcctgtgttcagagattaccagcagttc 433
Db 1348 ACCTCCAGTTCATGCCAAGCGTGAACGTTC 1379

RESULT 12
LOCUS I62451 1957 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 3 from patent US 5659123.
ACCESSION I62451
VERSION I62451.1 GI:2480399
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1957)
AUTHORS Van Rie,J., Jansens,S. and Peferoen,M.
TITLE Diabrotica toxins
JOURNAL Patent: US 5659123-A 3 19-AUG-1997;
FEATURES Location/Qualifiers
source
1..1957
/organism="unknown"
BASE COUNT 467 a 658 c 534 g 271 t 27 others
ORIGIN

Query Match 3.1%; Score 43.6; DB 5; Length 1957;
Best Local Similarity 48.8%; Pred. No. 5.1;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
Qy 699 ccaaggaagtgttcagttgggggagagagctacgggtcagagagcgggagagagact 758
Db 200 CCAAGGAGTGTATCAGAAGGCGATCAGCGTGGTGGCGACCTGCTGGCGGTGGGGCT 259
Qy 759 ttgctgcttcctgagctggctgcgtgctatggcatgccaggcatgagctccctgcagg 818
Db 260 TCCCTTCGGTGGTGGCTGTGAGCTTCTACACAACTTCTCTGAACACCATCTGGCCCA 319
Qy 819 accggcatggcgtaccatcttggttcagggggagatccttgacccttggtgacccaaggcg 878
Db 320 CGGAGGACCCCTGGAAGCCCTTCATGGAGCAGGTGGAGGCCCTGTGATGACCAAGATCG 379
Qy 879 gcaagtcccgcaaaagaaatccaagccacacagctgagtccttgaggacagagtggag 938
Db 380 CCGACTACGCCAAGCAAGGCCCTGGCCGAGCTGCAGGGGCTGCAGAACACCTGGAGG 439
Qy 939 ac 940
||

Db 440 AC 441
RESULT 13
GGU73678/c
LOCUS GGU73678.1 2657 bp DNA VRT 20-DEC-1996
DEFINITION Gallus gallus alpha-1,3-fucosyltransferase (CFT1) gene, complete cds.
ACCESSION U73678
VERSION U73678.1 GI:1657998
KEYWORDS chicken.
SOURCE Gallus gallus
ORGANISM Gallus gallus
REFERENCE 1 (bases 1 to 2657)
AUTHORS Lee,K.P., Carlson,L.M., Woodcock,J.B., Ramachandra,N., Schultze,L., Davis,T.A., Lowe,J.B., Thompson,C.B. and Larsen,R.D.
TITLE Molecular cloning and characterization of CFT1, a developmentally regulated avian alpha(1,3)-fucosyltransferase gene
JOURNAL J. Biol. Chem. 271 (51), 32960-32967 (1996)
MEDLINE 97115837
REFERENCE 2 (bases 1 to 2657)
AUTHORS Lee,K.P., Carlson,L.M. and Larsen,R.D.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1996) Immune Cell Biology Program, Naval Medical Research Institute, 8901 Wisconsin Ave., Bethesda, MD 20889, USA
FEATURES
source
1..2657
/organism="Gallus gallus"
/db_xref="taxon:9031"
170..1240
/gene="CFT1"
170..1240
/gene="CFT1"
/note="similar to mammalian fucosyltransferase FucTIV"
/codon_start=1
/product="alpha-1,3-fucosyltransferase"
/protein_id="AAC6060.1"
/db_xref="GI:1657998"
/translation="MELGPRMSPAARPGCPRRRRWALLGALGALALYVVCVELR
RRGSAGRGEVTVLLWMPFPPWRPADCRRTYNTGCLLSADRGRYGARAVLFH
HRDLALHGRGLPRGPPLPRORWVNFESPSHPLGLRGLAGLFWNTWSYRRSDV
FVPYLYEPSPRPPLPRKSLVAVVINSNNEEHARVYRKLKHLDDYVGAR
MALLEGSVTVSVSKYFLAFENSQHDYITEKLWKNAPASVVPVLGPRANERF
IPADSFTHVDDFPRLATYLFKLDKNKPSYRRYFAWRNKYIEHVTSFWDHEYCKVC
EAVRTAGNQLTKVNLGAFES"
BASE COUNT 581 a 712 c 818 g 546 t
ORIGIN

Query Match 3.1%; Score 43.6; DB 4; Length 2657;
Best Local Similarity 47.2%; Pred. No. 5.1;
Matches 133; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
Qy 149 tcagcccgccgagagagctgcgctgatactagccctctgcctggggccagcccaaa 208
Db 435 TCGGCGCTCAGAGGACAGCCGCTGATGTGTAGCGCCTCCCGGACGTGCGCGCGCGCCAG 376
Qy 209 caggagcactggcctggtcttcgcttcgcatgctcggtcttttcagctggtgcc 268
Db 375 GGGCGCGCGAAGGGCTCCACACAGACAGCCGCTACCTCGCCCTCGGGCGCGCGGCC 316
Qy 269 cgcgaggagctgcacgcatgccacctgcgcttttacacggcccccgctggccccgg 328
Db 315 GCCGAGCGCGCGCGCAGCTCCCGCACGAGAGCTACAGAGCCAGCGCGCGCGCCAGC 256
Qy 329 ctgcctctatgttgtagcatccgcgggttcggccctgggaccttgggggaaagtgg 388
Db 255 AGCGCGCCACAGCAGCGCGCGCCGACCTCCCGGCGAGCCCGCGCGCGCGCGCG 196
Qy 389 cagccggcgccgcccctgtgtcttgcaggagtagaccagcag 430
||


```
* 33747 34439: contig of 693 bp in length
* 34440 34449: gap of unknown length
* 34450 35117: contig of 668 bp in length
* 35118 35127: gap of unknown length
* 35128 35158: contig of 31 bp in length
* 35159 35168: gap of unknown length
* 35169 35199: contig of 31 bp in length
* 35200 35209: gap of unknown length
* 35210 35713: contig of 504 bp in length
* 35714 35723: gap of unknown length
* 35724 36549: contig of 826 bp in length
* 36550 36559: gap of unknown length
* 36560 37265: contig of 706 bp in length
* 37266 37275: gap of unknown length
* 37276 37957: contig of 682 bp in length
* 37958 37967: gap of unknown length
* 37968 39110: contig of 1143 bp in length
* 39111 39120: gap of unknown length
* 39121 39151: contig of 31 bp in length
* 39152 39672: contig of 511 bp in length
* 39673 39682: gap of unknown length
* 39683 40578: contig of 896 bp in length
* 40579 40588: gap of unknown length
* 40589 41266: contig of 678 bp in length
* 41267 41276: gap of unknown length
* 41277 41775: contig of 499 bp in length
* 41776 41785: gap of unknown length
* 41786 42470: contig of 685 bp in length
* 42471 42480: gap of unknown length
* 42481 43168: contig of 688 bp in length
* 43169 43178: gap of unknown length
* 43179 43674: contig of 496 bp in length
* 43675 43684: gap of unknown length
* 43685 44181: contig of 497 bp in length
* 44182 44191: gap of unknown length
* 44192 44704: contig of 513 bp in length
* 44705 44714: gap of unknown length
* 44715 44795: contig of 81 bp in length
* 44796 44805: gap of unknown length
* 44806 44894: contig of 89 bp in length
* 44895 44904: gap of unknown length
* 44905 45601: contig of 697 bp in length
* 45602 45611: gap of unknown length
* 45612 46297: contig of 686 bp in length
* 46298 46307: gap of unknown length
* 46308 46991: contig of 684 bp in length
* 46992 47001: gap of unknown length
* 47002 47513: contig of 512 bp in length
* 47514 47523: gap of unknown length
* 47524 48208: contig of 685 bp in length
* 48209 48218: gap of unknown length
* 48219 48916: contig of 698 bp in length
* 48917 48926: gap of unknown length
* 48927 48957: contig of 31 bp in length
* 48958 48967: gap of unknown length
* 48968 48998: contig of 31 bp in length
* 48999 49008: gap of unknown length
* 49009 49515: contig of 507 bp in length

Query Match      3.1%; Score 43.6; DB 44; Length 102713;
Best Local Similarity 50.5%; Pred. No. 5.1;
Matches 106; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 185 ccttgctggggccagcccccacagagagactggccttggtcttccttcgcttgcatg 244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72994 CCGGCGCCCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73053

Qy 245 tccggctcttttcagctgggtgcgcgcgagagctgcacgcacatgccacctgcgcttt 304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73054 CCGGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73113

Qy 305 tacacgccccgcctggccccgcgctgcgcctatgtttcttgagatccgcgcggttcgc 364
```

```
Db 73114 GGTGGCCCCCGCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 73173
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 365 cgtggacacttggtgggaaagtcgacgcg 394
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 73174 GGGCGCCCGCGTGGCGGGGGGTGGCGCGG 73203

RESULT 15
HS29K1 128779 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from PAC 29K1 on chromosome 6p21.3-22.2.
DEFINITION Contains glutathione peroxidase-like; zinc finger, ESTs, STS,
          tRNAs, olfactory receptor pseudogene.
ACCESSION Z98745
VERSION Z98745.1 GI:2924249
KEYWORDS 6p21.3-22.2; glutathione peroxidase-like; pseudogene; zinc finger.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 128779)
AUTHORS Tubby, B.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-1997) sanger.ac.uk/HGP/Chr6/ Sanger Centre,
          Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
          On Mar 4, 1998 this sequence version replaced gi:2511659.
          IMPORTANT: This sequence is the entire insert of clone 29K1. During
          sequence assembly data is compared from overlapping clones. Where
          differences are found these are annotated as variations together
          with a note of the overlapping clone name. Note that the variations
          annotated may not be found in the sequence submission corresponding
          to the overlapping clone as we submit sequences with only a small
          overlap as described above.
          This sequence was generated from part of bacterial clone contigs of
          human chromosome 6, constructed in collaboration by the Sanger
          Centre chromosome 6 mapping group and Armin Volz and Andreas
          Ziegler and David Ruddy.
          Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr6/ This sequence has been finished
          according to sequence map criteria as follows. An attempt is made
          to resolve all sequencing problems, such as compressions and
          repeats, but not necessarily within known annotated human repeat
          sequence elements (e.g. Alu). Where the sequence is ambiguous,
          there is an annotation using the 'unsure' feature key.
          The true left end of clone 29K1 is at 1 in this sequence. The true
          right end of clone 29K1 is at 128779.
          29K1 is from the library RPCI1 constructed at the Roswell Park
          Cancer Institute by the group of Pieter de Jong.
          For further details see http://bacpac.med.buffalo.edu/.
FEATURES
          Location/Qualifiers
            1..128779
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="6"
              /map="6p21.3-22.2"
              /clone="RPI-29K1"
              /clone_lib="RPCI-1"
              /note="2 copies of 60 mer 93 % conserved"
              /note="2 copies of 86 mer 96 % conserved"
              /note="2 copies of 167"
              /note="MIR repeat: matches 164. .70 of consensus"
              /note="MIR repeat: matches 262. .11 of consensus"
              /note="MIR repeat: matches 262. .11 of consensus"
              /note="match: 3' EST AA600215 clone 949990"
              /note="MIR repeat: matches 3. .142 of consensus"
              /note="MIR repeat: matches 3. .142 of consensus"
              /note="MIR repeat: matches 3. .142 of consensus"
              /note="AluSc repeat: matches 299. .1 of consensus"
```

```
misc_feature complement(7785..8029)
misc_feature /note="match: STS G31275"
repeat_region complement(7791..8027)
misc_feature /note="match: STS U50880"
gene 8155..8319
CDS /note="MER5B repeat: matches 1..177 of consensus"
gene complement(8542..8988)
CDS /note="match: STS U50881"
gene complement(9114..54893)
CDS /gene="dJ29K1.2"
CDS complement(join(9114..10396,11562..11706,16664..16995,
54120..54267,54519..54893))
/miscodon_start=1
/miscodon= "dJ29K1.2"
/miscodon_id="CAB11428.1"
/miscodon_xref="GI:2924250"
/miscodon_xref="SPTREMBL:O43724"
/miscodon_translation="WQGLLAVKVEEHEHSCGPESGLSRNPHTRIFRRRPFQFC
YQESQPRALQRLQELCHQWLPEMHTKEQILELLLEQFLTLPEELQAWVROHQP
VSGEAVILELELEDDPGEVLSHAHEQEEFVKKATPGAAQESSNQFOFLEQ
LQYNLREVCPOVIDGKAGTWVLPKREISOEVKSLIOVLGKQNGNITQIPEYGT
COREGLRKHQVSSVERPYTCSECGKFTONSILIEHQHTGKPYECDECGRAF
QSGLFQRLHTGKPYECSECGKFTONSILIEHQHTGKPYECDECGRAF
SVLKHQHTGKPYECSECGKFTONSILIEHQHTGKPYECDECGRAF
NMGKLSVOTLSTVTRHTGKPYECSECGKFTONSILIEHQHTGKPYECDECGRAF
KGGKVSQNSAFSTQPCWKTLSVDFVDDLEKHESEKCFGKQCKTLTHRSLALQH
QVYIAETCKCHGRKASVHNSLFRSTQMVAVLDIFLKHGTERPKYKNECOKA
SCQLALNHOHGFHAGEHCYCNCEGHFTQACFTISEVLEKHPCKLCKSKPKY
TRQMDLKNHTSREVFRQYPCFCYQTSQPREALSRLRELCHQWLPEMHTKEQ
ILELIVLQFQTLPELOAWVQHPSEGEVTVLEDERLEDFGEVHTHGEQ
EMFQETVRLKKEGSPNSLQSKAQKPYESPELQEQEDVETNEYGNLQVESEE
MEPHGKSSFEENDMSKARGETREPEITEEPSACREDKQDTQDENGSLTENSEE
HTEHQRCPEGSEGGDDCGKAFSQSHLIEHQHTGKPYECDECGKAFRTVLI
RKHILHGEKPYECDECGKAFSQSHLIEHQHTGKPYECDECGKAFRTVLI
IKIHTRPYCTQCNKFSRRSILTOHQVHTGKPYECDECGKAFRTVLI
IHKKEKCYQCKECSQSGLQHOIRHTGKPYECDECGKAFRTVLI
GERPKYCKGKAFRTVLIHQIRHTGKPYECDECGKAFRTVLIHQIRHTG
KPYQCECGKAFRTVLIHQIRHTGKPYECDECGKAFRTVLIHQIRHTG
VSV"
repeat_region 9508..9669
misc_feature /note="2 copies of 81 mer 82 % conserved"
repeat_region 10362..10833
repeat_region /note="match: STS G31315"
repeat_region 10746..10836
repeat_region /note="MER42c repeat: matches 1507..1414 of consensus"
repeat_region 10838..11135
repeat_region /note="AluJg repeat: matches 300..1 of consensus"
repeat_region 11142..11210
repeat_region /note="MER42c repeat: matches 1422..1356 of consensus"
repeat_region 12155..12414
repeat_region /note="AluJg repeat: matches 37..301 of consensus;
incomplete repeat"
repeat_region 12899..12939
repeat_region /note="MIR repeat: matches 75..115 of consensus"
repeat_region 13597..13894
repeat_region /note="AluSx repeat: matches 298..1 of consensus"
repeat_region 14680..14891
repeat_region /note="AluJg repeat: matches 300..82 of consensus;
incomplete repeat"
repeat_region 14904..15032
repeat_region /note="MIR2 repeat: matches 4..138 of consensus"
repeat_region 15265..15563
repeat_region /note="AluSx repeat: matches 300..1 of consensus"
repeat_region 15689..15945
repeat_region /note="MER20 repeat: matches 218..1 of consensus"
repeat_region 17296..17609
repeat_region /note="AluY repeat: matches 1..298 of consensus"
repeat_region 19815..20194
repeat_region /note="L1m2 repeat: matches 7..404 of consensus"
repeat_region 21612..21828
repeat_region /note="MLT1G repeat: matches 12..228 of consensus"
```

```
repeat_region 22148..22196
repeat_region /note="MLT1F repeat: matches 484..534 of consensus"
repeat_region 22480..23365
repeat_region /note="L1PA6 repeat: matches 890..1 of consensus"
repeat_region 23216..23746
repeat_region /note="L1 repeat: matches 5390..4862 of consensus"
repeat_region 23990..24291
repeat_region /note="AluSx repeat: matches 1..302 of consensus"
repeat_region 24421..24488
repeat_region /note="2 copies of 34 mer 90 % conserved"
repeat_region 25270..25422
repeat_region /note="FRAM repeat: matches 8..159 of consensus"
repeat_region 28847..28936
repeat_region /note="3 copies of 30 mer 92 % conserved"
repeat_region 28930..28964
repeat_region /note="U2 repeat: matches 1..35 of consensus"
repeat_region 29027..29194
repeat_region /note="AluJg repeat: matches 1..166 of consensus;
incomplete repeat"
repeat_region 29198..29730
repeat_region /note="L1MA3 repeat: matches 539..1 of consensus"
repeat_region 29585..30706
repeat_region /note="L1 repeat: matches 5390..4364 of consensus"
repeat_region 29946..30056
repeat_region /note="AluY repeat: matches 301..191 of consensus;
incomplete repeat"
repeat_region 30791..31101
repeat_region /note="AluJg repeat: matches 1..302 of consensus"
repeat_region 31113..31450
repeat_region /note="MLT1A1 repeat: matches 365..1 of consensus"
repeat_region 31564..31842
repeat_region /note="AluJg repeat: matches 1..300 of consensus"
repeat_region 32129..32413
repeat_region /note="AluSg repeat: matches 302..9 of consensus"
repeat_region 33231..33518
repeat_region /note="AluSg repeat: matches 301..1 of consensus"
repeat_region 35063..35099
repeat_region /note="MER5B repeat: matches 98..62 of consensus"
repeat_region 35431..35728
repeat_region /note="AluY repeat: matches 301..2 of consensus"
repeat_region 35894..35982
repeat_region /note="MIR repeat: matches 116..206 of consensus"
repeat_region 36291..36400
repeat_region /note="AluJg repeat: matches 1..111 of consensus;
incomplete repeat"
repeat_region 36458..36568
repeat_region /note="L1MA1 repeat: matches 1051..941 of consensus"
repeat_region 36575..36874
repeat_region /note="AluY repeat: matches 300..1 of consensus"
repeat_region 36877..37042
repeat_region /note="AluJg repeat: matches 300..135 of consensus;
incomplete repeat"
repeat_region 37502..37803
repeat_region /note="AluJg repeat: matches 2..302 of consensus"
repeat_region 39473..39896
repeat_region /note="MER42c repeat: matches 112..550 of consensus"
prim_transcript /note="match: 3' EST AA450090 clone 789109; Paired with
EST AA450147 matching this clone; match: 5' EST AA262022
clone IMAGE:686044; match: 5' EST AA450147 clone 789109;
Paired with EST AA450090 matching this clone"
repeat_region 42751..43982
repeat_region /note="TIGER1 repeat: matches 2418..1166 of consensus"
repeat_region 43987..44265
```

```
Query Match 3.1%; Score 43.6; DB 10; Length 128779;
Best Local Similarity 55.0%; Pred. No. 5.1;
Matches 110; Conservative 0; Mismatches 84; Indels 6; Gaps 1;
Qy 571 ctttgaagggccgctcgtgctgagccctgcagcagcagccgagccgagct 630
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54762 CTATCAGGAGTCCCTTGGGCGGCGGCGCTCTTCAAGACTCCAGGAGCTCTGCCATCA 54703
```

[illegible]

Search completed: May 24, 2000, 00:07:12
Job time: 11183 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:03:13 ; Search time 3267.04 Seconds

(without alignments)

652.577 Million cell updates/sec

Title: US-09-215-435-76

Perfect score: 526

Sequence: 1 ctgcctgctgctgtgtgcac.....cttctgtcaaaaaaaaaa 526

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*

- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_est45:*
- 80: gb_est46:*
- 81: gb_est47:*
- 82: gb_gss1:*
- 83: gb_gss2:*
- 84: gb_gss3:*
- 85: gb_gss4:*
- 86: em_gss1:*
- 87: em_gss2:*
- 88: em_gss3:*
- 89: em_gss4:*
- 90: gb_gss5:*
- 91: gb_gss6:*
- 92: gb_gss7:*
- 93: gb_gss8:*
- 94: gb_gss9:*
- 95: em_gss5:*
- 96: em_gss6:*
- 97: em_gss7:*
- 98: em_gss8:*
- 99: em_gss9:*
- 100: em_gss10:*
- 101: em_gss11:*
- 102: gb_gss10:*
- 103: gb_gss11:*
- 104: em_gss12:*
- 105: gb_gss12:*
- 106: gb_gss13:*
- 107: gb_gss14:*
- 108: gb_gss15:*
- 109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
C 1	441.4	83.9	44.3	79	AW303354	AW303354 xv18b01.x
C 2	114.8	21.8	40.9	84	B46318	B46318 HS-1063-B1-
C 3	64.6	12.3	22.3	74	AV381754	AV381754 AV381754
C 4	64.6	12.3	23.9	74	AV380943	AV380943 AV380943
C 5	63	12.0	17.8	74	AV379625	AV379625 AV379625
C 6	61.4	11.7	17.3	74	AV380189	AV380189 AV380189
C 7	51	9.7	4.6	107	AQ457768	AQ457768 HS_5142_A
C 8	48.8	9.3	23.7	23	R90989	R90989 yp93b05.r1
C 9	48.4	9.2	5.25	45	A1333042	A1333042 qq31h04.x
C 10	48.2	9.2	3.50	21	T74655	T74655 yc57e11.r1
C 11	48.2	9.2	4.13	33	A424642	A424642 zv91g02.r
C 12	48.2	9.2	4.27	24	H93888	H93888 yw08g08.r1
C 13	48.2	9.2	4.96	41	A1015730	A1015730 ot7ib08.s
C 14	48.2	9.2	5.06	41	A1065103	A1065103 HA0939 Hu
C 15	48.2	9.2	5.11	42	A110674	A110674 HAQ104 Hu
C 16	48.2	9.2	5.11	51	AF063548	AF063548 AF063548
C 17	48.2	9.2	5.21	43	A1207711	A1207711 HA3093 Hu
C 18	48.2	9.2	5.71	43	A1207590	A1207590 HA2954 Hu
C 19	48.2	9.2	5.83	45	A1336247	A1336247 qt45b07.x
C 20	48.2	9.2	6.44	60	A1807643	A1807643 wf49d04.x
C 21	48.2	9.2	6.62	41	A1065076	A1065076 HA0910 Hu
C 22	48.2	9.2	6.75	74	AW173259	AW173259 xj85c08.x
C 23	48	9.1	24.3	23	H47415	H47415 yu74e06.r1
C 24	46.8	8.9	4.02	24	H72811	H72811 yp07f10.r1
C 25	46.6	8.9	5.31	43	A1186874	A1186874 qe36b09.s
C 26	46.6	8.9	5.72	30	AA242899	AA242899 zif65b07.r
C 27	46.2	8.8	4.87	44	A1246161	A1246161 qt129b06.x
C 28	45.2	8.6	3.01	48	A1534767	A1534767 tn30e11.x
C 29	45.2	8.6	3.98	50	A1693185	A1693185 wd68e06.x
C 30	45.2	8.6	4.18	43	A1217375	A1217375 qh17g01.x
C 31	44.6	8.5	5.07	103	AQ193354	AQ193354 HS_3221_A
C 32	44.2	8.4	4.87	83	FR003770	FR003770 Fugu rubr
C 33	44.2	8.4	6.19	83	FR0037790	FR0037790 Fugu rubr
C 34	44	8.4	1.57	21	T74538	T74538 yc56c09.s1
C 35	43.8	8.3	5.11	80	AW385626	AW385626 PM0-LT003
C 36	43.4	8.3	4.95	33	AA424551	AA424551 zv91g02.s
C 37	43.4	8.3	5.40	37	AA705122	AA705122 zj195g09.s
C 38	42.8	8.1	11.01	82	CNS00HD9	CNS00HD9 Drosophila
C 39	42.2	8.0	4.59	80	AW385625	AW385625 PM0-LT003
C 40	42.2	8.0	5.03	80	AW385636	AW385636 PM0-LT003
C 41	42.2	8.0	5.13	80	AW385627	AW385627 PM0-LT003
C 42	42.2	8.0	5.23	80	AW385638	AW385638 PM0-LT003
C 43	42.2	8.0	5.43	80	AW385261	AW385261 RCO-LT003
C 44	42.2	8.0	5.46	80	AW385651	AW385651 PM0-LT003
C 45	42.2	8.0	5.57	80	AW385294	AW385294 RCO-LT000

ALIGNMENTS

RESULT	1
LOCUS	AW303354/c
DEFINITION	AW303354 443 bp mRNA EST 18-JAN-2000 xv15b01.x1 Soares_NFL_T_GBC.s1 Homo sapiens cDNA clone IMAGE:2813449 3', similar to SW:IBP_CARCR P00993 CHELONIANIN ;, mRNA sequence.

```

Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco.

FEATURES             Location/Qualifiers
     i.         443
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:2813449"
             /clone_lib="Soares_NFL_T_GBC_S1"
             /lab_host="DH10B"
             /note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
             a modified polylinker; Site.1: Not I; Site.2: Eco RI;
             Equal amounts of plasmid DNA from three normalized
             libraries (fetal lung NbHL19W, testis NHT, and B-cell
             NCI-GAP-GCB1) were mixed, and ss circles were made in
             vitro. Following HAP purification, this DNA was used as
             tracer in a subtractive hybridization reaction. The driver
             was PCR-amplified cDNAs from pools of 5,000 clones made
             from the same 3 libraries. The pools consisted of
             I.M.A.G.E. clones 297480-302087, 682632-687239,
             726408-728711, and 729096-731399. Subtraction by Bento
             Soares and M. Fatima Bonaldo."
     137 a         84 c         77 g         145 t
BASE COUNT
          100

```

RESULT 2
B46318/c
GLOCUS
DEFINITION B46318-1063-B1-A10-MR.abi DNA
GSS
CIT Human Genomic Sperm Library C Homo sapiens-genomic clone Plate-Ct796 Col=19 Row=B genomic survey sequence.
B46318
sequence.

ACCESSION B46318.1 GI:2551152
VERSION B46318.1
KEYWORDS GSS.

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 173)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,T., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
Kai,C., Kawai,Y., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Takahashi,F., Tateo,M., Tomimaga,N.,
Tsukuda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno.H., et al.)
Unpublished (1999)
On Jun 22, 1998 this sequence version replaced gi:3247200.
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel.: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsuyama,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

```

Location/Qualifiers
1. .173
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/cloned="9230021C08"
/clone_lib="RIKEN full-length enriched, adult male
epididymis"
/sex="male"
/tissue_type="epididymis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site.1: SalI; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATTAATCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites...
vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"

```

```

BASE COUNT      45 a   33 c   31 g   64 t
ORIGIN

Query Match
Best Local Similarity 69.08; Pred. No. 2.6e-05;
Matches 100; Conservative 0; Mismatches 41; Indels 4; Gaps 1;

Qy 338 gaagtgtctgtgcacatccgaaataaagacacagaagaattcagactgatttggaaa 397
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20 GAGGTTTCGGATGGCCATCTGGAATAAAGATTCAAGAAGATTACAGACCAATCTTTAGG 79
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 398 tctttgtaataattccataatgttttaagcttcacatattgttgctatttctctgacccta 457
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 TTTTGGTTCGTGCCAATATCTTTGACACT- - -TATATTCACTATTTCCTCTCTCCTA 135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 458 gttttgtcttcctggaaattaaact 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 ATTTTCTTCTTCCTGGAATAAATAAT 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
LOCUS      AQ457768      436 bp      DNA      GSS      23-APR-1999
DEFINITION HS_5142_A1_E12.SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate-718 Col-23 Row-I, genomic survey sequence.
ACCESSION  AQ457768
VERSION     AQ457768.1 GI:4636408
KEYWORDS   GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 436)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 718 row: I column: 23
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 436.
FEATURES
    source
        1. .436
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate-718 Col-23 Row-I"
                /clone_lib="RPCI-11 Human Male BAC Library"
                /sex="male"
                /note="Vector: pBACE3.6; Genomic sequence of BAC ends"
BASE COUNT      150 a   77 c   82 g   122 t   5 Others
ORIGIN

Query Match
Best Local Similarity 9.7%; Score 51; DB 107; Length 436;
Matches 99; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```

```

Qy 131 taaaagatccctgcacaaattggacatgaatttggagagctgctatgaagtccacttagat 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 TCACAGAACTACTGCACTATCTCCGGAAGGACCTGTAATTCATATTTGACCAAGAT 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 191 attctacaacagacacccccaagatgtgaaacttttctctccggctgtaagtgcga 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 TCTACTACAACACTTTGACATTTTATATGTGAGCCCTTTGTCTTCAGTGGCTGTGGAGGCA 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 251 accttaacaactcaagcttaaaatagaacgtgaagtagcctgtgttgcaaaaatacaaa 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 ACAGAAACAACCTTTAAACACAGAAATATTTCTGTGAAAAAATGTGTATTACTCAAAAGTAA 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
LOCUS      R90989      237 bp      mRNA      EST      25-AUG-1995
DEFINITION YP93b05.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
            IMAGE:194961 5' similar to gb:X04225 ALPHA-1-MICROGLOBULIN
            (HUMAN);, mRNA sequence.
ACCESSION  R90989
VERSION     R90989.1 GI:958529
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 237)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
TITLE      The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     On Sep 21, 1992 this sequence version replaced gi:276046.
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 755
            High quality sequence starts: 1
            High quality sequence stops: 1
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Seq primer: M13RPI
            High quality sequence stop: 1.
FEATURES
    source
        1. .237
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="GDB:3764011"
                /db_xref="taxon:9606"
                /clone="IMAGE:194961"
                /clone_lib="Soares fetal liver spleen INFLS"
                /sex="male"
                /dev_stage="20 week-post conception fetus"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                1st strand cDNA was primed with a Pac I - oligo(dT) primer
                [5' AACGGAAGAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Pac I and cloned into the Pac I
                and Eco RI sites of the modified pT73 vector. Library
                went through one round of normalization. Library
                constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      62 a   56 c   69 g   48 t   2 others
ORIGIN

```

Query Match	9.3%	Score 48.8	DB 23	Length 237
Best Local Similarity	57.7%	Pred. No. 0.033		
Matches	86	Conservative	0	Mismatches 63
				Indels 0
				Gaps 0
QY	134	agatccctgcgaatggacatgaatttggagctgctatgaagttcaccttttagatatt	193	
Db	67	AGATTCTCTGCCAGCTGGGCTACTGCGCGGTCCTGCATGGGNAATGACCAAGTAGGTATT	126	
QY	194	tctacaacagaaacctccaaagaatgtgaaactttgtctcttcgcgctgttaatggcaacc	253	
Db	127	TCFATAATGTTACATCATCGCTGTGAGACTTCCAGTAGCGGNGCTGCATGGCAACG	186	
QY	254	ttacaaccttcaagcttaaatagaacgt	282	
Db	187	GTAAACAACCTTCGCACAGAAAGGAGGTGT	215	

RESULT	9	
AI1333042/c		
LOCUS	525 bp	mrna
DEFINITION	AI1333042	EST
	qq31h04.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1934167 3' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN);, mRNA sequence.	
ACCESSION	AI1333042	
VERSION	AI1333042.1	GI:40696501
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 525)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	On Jan 17, 1998 this sequence version replaced gi:2043458.	

real: 13027 420 13200
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1327 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 410.

```

FEATURES
source
    1. 525
    Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:1934167"
        /clone_lib="Soares_NHMPU_S1"
        /tissue_type="Pooled human melanocyte, fetal heart, and
        pregnant uterus"
        /lab_host="DH108"
        /note="Organ: mixed (see below); Vector: pT73D-Pac
        (Pharmacia) with a modified polylinker; Site.1: Not I;
        Site.2: Eco RI; Equal amounts of plasmid DNA from three
        normalized libraries (melanocyte 2NbHM, pregnant uterus
        NBHPU, and fetal heart NBHH19W) were mixed, and ss circles
        were made in vitro. Following HAP purification, this DNA
        was used as tracer in a subtractive hybridization
        reaction. The driver was PCR-amplified cDNAs from pools
        5,000 clones made from the same 3 libraries. The pools
        consisted of I.M.A.G.E. clones 260232-265223,
        340488-345479, and 484488-489479."
    115 a 147 c 131 g 132 t
BASE COUNT

```

Query Match 9.2%; Score 48.4; DB 45; Length 525;
Best Local Similarity 60.8%; Pred. No. 0.041;
Matches 79; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy	134	aagatccctgcaaaattggacatgaattttggaagtctgatgaagtccaatttagatatatt	193
Db	510	AAGATTCTCCAGCTGGGCTACTCGCGGTCCCTGCATTTGAATGCCACGACGGTATT	451
Qy	194	tctacaagagaacctccaaagatgtgaacttttgccttcocgggtgtaatggcaacc	253
Db	450	TCATATAATGTFATCATCCATGGCCTGTGAGACTTCCAGTACGGGGGGTGTCATGGCAACG	391
Qy	254	ttaacaactt	263
Db	390	GTAACAACCTT	381

RESULT 10

LOCUS	T74655	350 bp	mrna	EST	02-MAR-1995
DEFINITION	yc57ell.r1 Stratagene liver (#937224) Homo sapiens cDNA clone IMAGE:84812 5' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN);, mRNA sequence.				

ACCESSION T74655
VERSION T74655.1 GI:691330
KEYWORDS EST.
SOURCE human.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

REFERENCE
AUTHORS

1 (bases 1 to 350).
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, M., DuBuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, J.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,
and Marta, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE
97044478
Contact: Wilson RK
Washingtton University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 62
COMMENT

Insert Size: 0.2
High quality sequence stops: 281 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 62 Std Error: 0.00
Seq primer: M13RP1

FEATURES	High quality sequence stop: 281.
source	Location/Qualifiers
	1. .350

```
/organism="Homo sapiens"  
/db_xref="GDB:501869"  
/db_xref="taxon:9606"  
/clone_1="IMAGE:84811"  
/clone_lib="Stratagene liver (#937224)"  
/sex="male"  
/dev_stage="49 years old"  
/lab_host="SOLR cells (kanamycin resistant)"  
/note="Organ: liver; Vector: pBluescript SK+;  
EcoRI; Site_2: XhoI; Cloned unidirectionally  
into dt. Hepatectomy from normal male caucasian  
adult; Insert size: 1.1 kb; Uni-ZAP XR Vector; -5'  
sequence: 5' GAATTCGGCAGCGAG 3' -3' adaptor s  
CTCCGACCTTTTCTTTTTTTTTTTTTTTT 3'"
```

BASE COUNT	ORIGIN	78 a	92 c	103 g	76 t	1 others
		78 a	92 c	103 g	76 t	1 others

Query Match	9.2%;	Score 48.2;	DB 21;	Length 350;
Best Local Similarity	57.7%;	Pred. No. 0.046;		

Matches	86;	Conservative	0;	Mismatches	63;	Indels	0;	Gaps	0;
Qy	134	aagatccctgc	aaatttgga	catgaatttgg	agctgctatgaag	tctcaatttagatatt	193		
Db	36	AAAGATCTCC	AGCTGGCTACTCG	CGCGGTCCTTCG	ATGGATGAC	CACGACGGTATT	95		
Qy	194	tctacaaga	caactccaaa	agatgtg	aaacttttgtcttcc	cgctgtaatgg	caacc	253	
Db	96	TCTATAATG	TACATCCATGG	CGCTGTGAG	ACTTTCAGTAC	GGCGGTCGTCATGG	GCACG	155	
Qy	254	ttacaactt	caagctt	aaataag	aacgt	282			
Db	156	GTAACAACT	TCGTCAG	AAAGAG	TGT	184			
RESULT	11								
AA424642									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
FEATURES									
source									
BASE COUNT									
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
Qy	134	aagatccctgc	aaatttgga	catgaatttgg	agctgctatgaag	tctcaatttagatatt	193		

QY 134 aagatccctgcaaatggaacatattttggaagctgctatgaagtccaacttagatatt 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 170 AAGATTCCTGCAGCTGCTCGCGGCTCCCTGCATGGGAATGACCACAGGTATT 229
QY 194 ttacaacagaacctccaaagatgtgaaactttgtcttctccggctgtaagtgcaccc 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 230 TCTATAATGGTACATCCATGGCTGTGAGACTTTCAGTACGCGCGCTGCATGGGCAACG 289
QY 254 ttacaactctcaagcttaaaatagaaactgaagtacc 291
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 GTAACAACCTTCGTACAGAAAGGAGTGTCTGNCAGAC 327

RESULT 13
AI015730/c
LOCUS AI015730 496 bp mRNA EST 27-AUG-1998
DEFINITION ot71b08.sl Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:1622199 3' similar to gb:X04225 ALPHA-1-MICROGLOBULIN
(HUMAN);, mRNA sequence.
ACCESSION AI015730
VERSION AI015730.1 GI:3230066
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1406864.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 2157 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence status: 272.

FEATURES
source
1..496
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1622199"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCTTAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 105 a 145 c 128 g 118 t
ORIGIN

Query Match 9.2%; Score 48.2; DB 41; Length 496;
Best Local Similarity 57.7%; Pred. No. 0.046;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggaacatattttggaagctgctatgaagtccaacttagatatt 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 AAGATTCCTGCAGCTGCGGCTACTCGCGGTTCTGCTGATGGGAATGACCACAGGTATT 433
QY 194 ttacaacagaacctccaaagatgtgaaactttgtcttctccggctgtaagtgcaccc 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 TCTATAATGGTACATCCATGGCTGTGAGACTTTCAGTACGCGCGCTGCATGGGCAACG 373

QY 254 ttacaactctcaagcttaaaatagaaactgaagt 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 GTAACAACCTTCGTACAGAAAGGAGTGT 344

RESULT 14
AI065103
LOCUS AI065103 506 bp mRNA EST 11-NOV-1999
DEFINITION HA0939 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
sequence.
ACCESSION AI065103
VERSION AI065103.1 GI:6359375
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 506)
AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Li, W., Wu, J., Zhou, S., Liu, M.
and He, F.
TITLE Expression profile analysis of a human fetal liver cDNA library
JOURNAL Unpublished (1998)
COMMENT On Jan 9, 1998 this sequence version replaced gi:937706.
Contact: Yongtao Yu
Department of Hematology
Beijing Institute of Radiation Medicine
27 Raiping Road, Beijing 100850, P.R.China
Tel: 0086-10-682159479
Fax: 0086-10-68214653
Email: Yyt48@yahoo.com.

FEATURES
source
1..506
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal liver cDNA library"
/tissue_type="liver"
/dev_stage="fetal"
/lab_host="MC1061/P3"
/note="Vector: pCDNAL"
BASE COUNT 116 a 136 c 147 g 107 t
ORIGIN

Query Match 9.2%; Score 48.2; DB 41; Length 506;
Best Local Similarity 57.7%; Pred. No. 0.046;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggaacatattttggaagctgctatgaagtccaacttagatatt 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20 AAGATTCCTGCAGCTGGGCTACTCGCGGCTCCCTGCTGATGGGAATGACCACAGGTATT 79
QY 194 ttacaacagaacctccaaagatgtgaaactttgtcttctccggctgtaagtgcaccc 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 TCTATAATGGTACATCCATCCATGGCTGTGAGACTTTCAGTACGCGCGCTGCATGGGCAACG 139
QY 254 ttacaactctcaagcttaaaatagaaactgaagt 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140 GTAACAACCTTCGTACAGAAAGGAGTGT 168

RESULT 15
AI110674
LOCUS AI110674 511 bp mRNA EST 11-NOV-1999
DEFINITION HA0104 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
sequence.
ACCESSION AI110674
VERSION AI110674.1 GI:6359539
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 511)
 AUTHORS Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M. and He,F.
 TITLE Expression profile analysis of a human fetal liver cDNA library
 JOURNAL Unpublished (1998)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397605.
 Contact: Yongtao Yu
 Department of Hematology
 Beijing Institute of Radiation Medicine
 27 Taiping Road, Beijing 100850, P.R.China
 Tel: 0086-10-68159479
 Fax: 0086-10-68214653
 Email: yyt48@yahoo.com.

FEATURES
 source
 1..511
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human fetal liver cDNA library"
 /tissue_type="liver"
 /dev_stage="fetal"
 /lab_host="MC1061/p3"
 /note="Vector: pCDNA1"

BASE COUNT 114 a 140 c 155 g 102 t
 ORIGIN

Query Match 9.2%; Score 48.2; DB 42; Length 511;
 Best Local Similarity 57.7%; Pred. No. 0.046;
 Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 134 aagatccctgcaaatggacatgaatttgggaagctgctatgaagttcacttttagatatt 193
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 215 AAGATTCCTGCCAGCTGGGCTACTCGGCCGCTCCCTGCATGGGAATGACCAGCAGGTATT 274
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 194 tctacaacagaaacctccaagaagtgtgaaacttttcttccgcgctgtaatggcaacc 253
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 275 TCTATAATGGTACATCCATGGCCCTGTGAGACTTTCACAGTAGCGGGCTGCATGGGCAACG 334
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 254 ttaacaacttcaagcttaaaatagaacgt 282
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 335 GTAACAACCTTCGTCACACAGAAAAGGAGTGT 363

Search completed: May 23, 2000, 22:03:18
 Job time: 3956 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2000, 13:28:15 ; Search time 3022.29 Seconds
(without alignments)
677.106 Million cell updates/sec

Title: US-09-215-435-78
Perfect score: 542
Sequence: 1 cacgacctgtggccatgat.....aaaaaaaaaaaaaaaaaaaaa 542

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

- 45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	421	77.7	491	41	AI018488	AI018488 ov63f08.s
C 2	44	8.1	518	45	AI394648	AI394648 tg08g04.x


```

ACCESSION   AI457661
VERSION     AI457661.1 GI:4310530
KEYWORDS    EST.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE   1 (bases 1 to 567)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138424.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2833 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 60.

FEATURES             source
    Location/Qualifiers
        1..567
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2144764"
            /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
            /lab_host="DH10B"
            /note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
            a modified polylinker; Site.1: Not I; Site.2: Eco RI;
            Equal amounts of plasmid DNA from five normalized
            libraries were mixed, and ss circles were made in vitro.
            Following HAP purification, this DNA was used as tracer in
            a subtractive hybridization reaction. The driver was
            PCR-amplified cDNAs from pools of 5,000 clones made from
            the same 5 libraries. The pools consisted of the following
            libraries and clones: Soares NBHSF pool 1:
            309384-310919, 323208-325895 Soares NB2HP pool 1:
            145032-147335, 147720-148103, 148872-149255, 15002 -
            150407, 151176-152327 Soares NB2HF8-9W pool 1:
            758280-760583, 772104-774407 Soares NBHPA pool 1:
            304776-306311, 320136-322823, 326280-326663 Soares NBHOT
            pool 1: 723720-726407, 739080-740999 Subtraction by Bento
            Soares and M. Fatima Bonaldo."
            134 a 152 c 133 g 147 t 1 others

BASE COUNT      134 a 152 c 133 g 147 t 1 others

Query Match      7.9%; Score 42.8; DB 46; Length 567;
Best Local Similarity 61.8%; Pred. NO. 2.3;
Matches 68; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 433 ccagagctgtgttcaccctgttcccgagcgtccaccatgagtgaggaggaggagg 492
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 CCAGCATCCCTATTTCCTCCGAGCGGAAAAAGCGCAGGAGGACCGAGGAGGTGTG 82

QY 493 agtgattgaataaagagcttttcaatgaaataaaataaaataaaataaa 542
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 GTTAACCTATATAAAAGGTTTTTTTTTTTAAAAAATAAAATAAAATAAA 32

RESULT 5
AW129541/c
LOCUS      AW129541 653 bp mRNA EST 25-OCT-1999
DEFINITION xel18f08.x1 NCI-CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2607495 3',
            similar to TR:Q60525 Q60525 GUANINE NUCLEOTIDE-BINDING PROTEIN BETA
            5. ; contains MER22.b3 MER22 MER22 repetitive element ;, mRNA
            sequence.
ACCESSION  AW129541
VERSION     AW129541.1 GI:6117485
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens

```

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 653)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Dec 20, 1995 this sequence version replaced gi:1135916.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 322.

FEATURES

source

Location/Qualifiers
1..653
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2607495"
/clone_lib="NCI-CGAP_Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.48 kb. Life technologies catalog #:
113422016"

BASE COUNT 156 a 182 c 156 g 158 t 1 others
ORIGIN

Query Match 7.9%; Score 42.8; DB 69; Length 653;
Best Local Similarity 75.7%; Pred. NO. 2.3;
Matches 53; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 473 atgagtggaggaggaggaggaggaggaggaggaggaggaggaggaggaggagg 532
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 AGGACCGGAGGAGGTGTGGTGAATAAATAAAGGTTTTTTTGTAAAAA 31

QY 533 aaaaaaaaaa 542
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30 AAAAAAAAAA 21

RESULT

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AI440449/c 382 bp mRNA EST 13-APR-1999
tj03c08.x1 NCI-CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2140430 3',
similar to TR:O15195 O15195 VILLIN-LIKE PROTEIN. ;, mRNA sequence.
AI440449
AI440449.1 GI:4300306
EST.
human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 382)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jun 5, 1998 this sequence version replaced gi:3188250.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Woskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 DNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CQAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio-llnl.gov/brbp/image/image.html

Insert Length: 2532 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 258.

FEATURES
source

```

Location/Qualifiers
1. 382
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:2140430"
  /clone_11b="NCI-CGAP_Gas4"
  /tissue_type="poorly differentiated adenocarcinoma with
    signet ring cell features"
  /lab_host="DH10B"
  /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
    Site_2: NotI; Cloned unidirectionally. Primer: Oligo 47.
    Average insert size 1.69 Kb. Life Technologies catalog #:
```

BASE COUNT	81 a	88 c	91 g	121 t	1 others
ORIGIN					

Query Match 7.8%; Score 42.4; DB 46; Length 382;
Best Local Similarity 51.0%; Pred. No. 2.8;
Matches 100; Conservative 0; Mismatches 96; Indels 0

Qy 347 gaaggcagaagtggctaagaaaatgttcttcttagtgcctccttcttctgctgcctcct 406

db 216 GGAGGCGCGGCGGAAAAAACAAGTGGGGTTTTTTTGAACCAAGCCCTTTGACTG 157

Qy 407 cctcctccacctgtctctcctccctacccagagctctgtgttcaacctgttccccagagcc 466

Db 156 CCCCTATCCCCCTGGACCCCCAACATACCTACAAATGTTGGGGGGGGCCCTGCTTCCCATCCCC 97

Qy 467 tccaccatgagtggaggggaagtgggggagtgttgaataaagagcttttccaatgaaaa 526

Db 96 TCAAGGCTTTTGGTCATCTCTTTGGGTGTGAGTAAAGCAGGCAGCCCATAAAAA 37

Qy 527 aaaaaaaaaaaaaa 542
 |||||
 Db 36 aaaaaaaaaaaaaa 21

Db 36 AAAAAAAAAAAAAA 21
| | | | | | | | | | | | | | | |

RESULT 7

LOCUS	AI538908	376 bp	mRNA	EST	13-MAY-1999
DEFINITION	tp99g10.x1 NCI-CCAP-J73 Homo sapiens cDNA clone IMAGE:2204610 3' similar to contains element OFR repetitive element ; , mRNA sequence.				

ACCESSION AI538908
VERSION AI538908.1 GI:4453043
KEYWORDS EST.

SOURCE	ORGANISM	human.	Homo sapiens

REFERENCE 1 (bases 1 to 376)

Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3188659.
Contact: Robert Strauchman Ph.D.

CONTACT: ROBERT STRAUSBERG, F.N.D.
Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome
Center
Clone distribution: NCI-CGAP clone distribution
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnlnl.gov/dbbrp/image/image.html

Insert Length: 2456 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 360

FEATURES

```

location/Qualifiers
1..376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2204610"
/clone_lib="NCI-CCAP_Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:
Site_2: NotI; Cloned unidirectionally. Primer:
Average insert size 1.45 kb. Life Technologies

```

	11341-010		
BASE COUNT	113 a	77 c	80 g 106 t
ORIGIN			

Query Match 7.7%; Score 42; DB 47; Length 376;
Best Local Similarity 55.5%; Pred. No. 3.4;
Matches 81; Conservative 0; Mismatches 65; Indels

Qy 397 gctgcctcctcctcctccacctgctctcctcctccctaccagagctctgtgttcacctgtt 456

D_b 176 GTTGGGCGCTGGCCCGTAGTTTCATCCCGTCCCGCGCTTTGGGGGCGCGTGTTCGCGCGCGG 117

Qy 457 cccagagcctcaaccatgagtggagggaagtggggagtgattgaaataaagagcctttt 516

db 116 GGGCGTTCCTCCCGCTGCCCTACCTGGGTAAAGGGCGTAAATGGGAATAAAATTCCTTTT 57

Qy 517 caatgaaaaaaaaaaaaa 542

Db 56 CCCTAAAAA 31

RESULT 8

AI411581/C	AI411581	353 bp	EST	09-FEB-1999
LOCUS	EST239875	Normalized rat ovary,	Bento Soares	Rattus sp. cDNA clone
DEFINITION	ROVET35 3' end, mRNA sequence.			

ACCESSION AI411581
VERSION AI411581.1 GI:4255085
KEYWORDS EST.

SOURCE	ORGANISM	Rattus sp.
		Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 353)

AUTHORS
Lee, N.H.,
Kerlavage.

TITLE	Rat Genome Project: Generation of a Rat EST (REST) C
JOURNAL	Gene Index Unpublished (1998)

```

FEATURES
  source
    Location/Qualifiers
      1. 353
      /organism="Rattus sp."
      /db_xref="taxon:10118"
      /clone="ROVET35"
      /clone_lib="Normalized rat ovary, Bento Soares"
      /note="Organ: ovary; Vector: pT73Pac; Site_1: EcoRI;
      Site_2: NotI"
      68 a 90 c 110 g 85 t
BASE COUNT
ORIGIN

Query Match
  Best Local Similarity 7.7%; Score 41.6; DB 46; Length 353;
  Matches 92; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 113 tggagctcaaggagttgttcacgcgaacaggactgcgagactgctgctcctccaaactg 172
Db 346 TGGAGGACGGTGAGATCTGCGTAACAGTATGCAGTGTGAAGCAGATGCTGCCAACATG 287
QY 173 ctccagacaattgcgagtcgactgcgcggaaggagggtccgagggcagctgtgtcaca 232
Db 286 ACACCATCTCGGCATCGCCCGTGCACACACAAGGCCATGGAACAGCGAGTCTCCC 227
QY 233 cgcaggtgttcttggcccaatagagctgtcctgctgcctgcggaacctgactgtgt 288
Db 226 CAAAGACCTCTATGGATGTACTACAGTGTCCCTGTGAGCGGGGCTGACTTGT 171

RESULT 9
AI679266/c
LOCUS
DEFINITION
  tu62e02.x1 NCI-CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2255642 3',
  similar to SW:PEPC_HUMAN P20142 PROGASTRICUSIN PRECURSOR ;, mRNA
  sequence.
ACCESSION
  AI679266
VERSION
  AI679266.1 GI:4889448
KEYWORDS
  human.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Carnivora; Hominidae; Homo.
  1 (bases 1 to 727)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  On Mar 10, 1998 this sequence version replaced gi:2948628.
  Contact: Robert Strausberg, Ph.D.
  Tel: (301) 496-1550
  Email: Robert.Strausberg@nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: Life Technologies, Inc.
  DNA Sequencing by: Greg Lennon, Ph.D.
  Cloned through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 205.
Location/Qualifiers
  1. 727
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:2255642"
  /clone_lib="NCI-CGAP_Gas4"
  /tissue_type="poorly differentiated adenocarcinoma with
  signed ring cell features"
  /lab_host="DH10B"
  /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
  Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

```

```

Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"
BASE COUNT 133 a 197 c 206 g 186 t 5 others
ORIGIN

Query Match
  Best Local Similarity 7.6%; Score 41.4; DB 50; Length 727;
  Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 404 cctcctcctccacctgctcctcctccaccagagctgtgttcacccctgttcccccaga 463
Db 194 CCCCCTTTCCCTTTGGCCCGCCCGCCCGCCCGGGGGGGTGTCTTTTCCCGCCCTGGG 135
QY 464 gctccaccatgagtgagggagtgaggagtgatgaataaagagagctttttcaatgaa 523
Db 134 ATCCCCCCCCCTTTTGGGGCCCGGGCCCTTTCTTAAAAAAAAGGTTTTTTTATA 75
QY 524 aaaaaaiaaaaaaiaaaaaa 542
Db 74 AAAAAAAAAAAAAAAAAA 56

RESULT 10
AI593765/c
LOCUS
DEFINITION
  v163a09.x1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
  clone IMAGE:976888 3', mRNA sequence.
ACCESSION
  AI593765
VERSION
  AI593765.1 GI:4602813
KEYWORDS
  EST.
SOURCE
  house mouse.
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 178)
  Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
  Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
  Ritter, E., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
  Waterston, R., and Wilson, R.
  The WashU-NCI Mouse EST Project 1999
  Unpublished (1999)
  On May 18, 1998 this sequence version replaced gi:3137509.
  Contact: Marra M/WashU-NCI Mouse EST Project 1999
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@wustl.edu
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:557616
  This clone was previously sequenced on the 5' end only, this new
  data is from the 3' end
  Possible reversed clone: polyt not found
  Seq primer: Primer name ambiguous.
  Location/Qualifiers
    1. 178
    /organism="Mus musculus"
    /strain="B6D2 F1/J"
    /db_xref="taxon:10090"
    /clone="IMAGE:976888"
    /clone_lib="Knowles Solter mouse blastocyst B1"
    /tissue_type="blastocyst"
    /dev_stage="embryo (pre-implantation)"
    /lab_host="DH10B"
    /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally from mRNA prepared
    from 800 blastocysts. Primer: SalI(dt);
    5'-CGTCGACCGTCGACCGCTTTTGT-3'. cDNAs were
    cloned into the NotI/SalI sites of a pSPORT vector (Life
    Technologies). Two different size selections: B1 (larger

```


Search completed: April 30, 2000, 13:28:20
Job time: 10680 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:03:23 ; Search time 3267.04 Seconds
(without alignments)
1107.892 Million cell updates/sec

Title: US-09-215-435-113
Perfect score: 893
Sequence: 1 gccatgacgcgcgtcac.....aaaaaaaaaaaaaaaa 893

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query


```

Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:996809
Seq primer: custom primer used
High quality sequence stop: 503.
Location/Qualifiers
1. .543
/organism="Mus musculus"
/strain="C57Bl."
/db_xref="taxon:10090"
/clone="IMAGE:2099877"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH108"
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACTCTGG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dr) primer
[ATGCGCCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
primer CGACCTGTCGACGTCGACACA."
121 a 167 g 94 t

```

Query Match	43.7%	Score	390.4;	DB	62;	Length	543;
Best Local Similarity	91.0%;	Pred.	No. 4.4e-93;				
Matches	415;	Conservative	0;	Mismatches	41;	Indels	0;
Gaps	0;						
QY	1	gcacatcacgacgacgtccaccctgggaagatgccctgtccaacgtggacctgttgaagag	60				
Db	88	GCCATGACCACCCACGTCACTTGGAAAGTCGCCGTGTCCAACGTGGACCTGCTCGAGGA	147				
QY	61	cttcctcctccccaccagcgccatgcacgtcagcctccaacttcccattcctcatcatgaccag	120				
Db	148	CTGCCCTCCACACCAGCAGCATGTATGAGCCCCGCCATCCTCTCATCATGTACCAG	207				
QY	121	gctaactttgacaaacttgtagcacagaatacttatctcaggccattgcaagtac	180				
Db	208	GCTAACTTTGACAGAACTTCGAGGACAGAATGCATTGTACGGGCATTGCAAGGTAC	267				
QY	181	attgagcaggctacagtccactccagcatgaatgagatgctggagggaagacatgagtat	240				
Db	268	ATTGACGAGGCAACGGTGCCACTCCAGCATGAATGAAATGCTGGAGGAGGGCCACGACTAT	327				
QY	241	gcsgtcatcgttacacctggcagcagctgttccgggceattcccacaggtaaatgcaac	300				
Db	328	GCASTCATGCTGTACCTGGCGAGCTGTCTCCGGGCCATCCGCGAGGTGAAGTGGCAAT	387				
QY	301	gagcagcccaaccgagtagagatctatgagaagacagtagaggtgctggagccggaggtc	360				
Db	388	GAGCAGGCCAAATCCGGTTCGAGATCTACGAGAAGACGGTGGAGGTGCTGGAGCCAGAGGTC	447				
QY	361	accaagctcatgaattcatgtattttcagcgaagccatcgagcggttctgcagcgag	420				
Db	448	ACCAAGCTCATGAAGTTTCATGTACTTTCACGGCAAGGCCATTCAGCGCTTCTGTAGCGAG	507				
QY	421	gtgaagcggctgtgccatgccgagcgaggaagac	456				
Db	508	GTGAAGCGGCTGTGTCAACGTCAGCGTAGCGCAGGAAGCAC	543				

RESULT	4
AI405710	
LOCUS	699 bp mRNA EST 08-FEB-1999
DEFINITION	GH25824.5prime GH Drosophila melanogaster head pot2 Drosophila melanogaster cDNA clone GH25824 5prime, mRNA sequence.
ACCESSION	AI405710
VERSION	AI405710.1 GI:4248797
KEYWORDS	EST.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophylla.
REFERENCE	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M. BDGP/HMMI Drosophila EST Project Unpublished (1997)
TITLE	On May 7, 1998 this sequence version replaced gi:3119095.
JOURNAL	Contact: Harvey, D.
COMMENT	G. M. Rubin-Molecular and Cell Biology University of California Berkeley 539 LSA, Berkeley, CA 94720-3200, USA Fax: 510 643 9947 Email: http://www.fruitfly.org/EST , est@fruitfly.berkeley.edu Plate: 258 row: B column: 12 High quality sequence stop: 524.

FEATURES
SOURCE

```

1. 1099
source
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH25824"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/notes="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
167 a 197 c 198 q 137 t
BASE COUNT

```

BASE COUNT	167	A	197	C	198	G	137	T	ORIGIN
Query Match	43.0%	Score	384.2	DB	46	Length	699		
Best Local Similarity	75.2%	Pred. No.	2.1e-91						
Matches	479	Conservative	0	Mismatches	158	Indels	0	Gaps	0
QY	3	catgtatcacgcagctcaccctggaagatgcccctgtccaaacgtggaccctgcttgaagagct	62						
DB	63	CATCGGGAAGATTACGCTAGCCGACGGCGCTGCCAACGCTGGAGGTGTTGGACGAGCT	122						
QY	63	tccctcccgaccagcagccatgcacgcagctcatcgagctccaccctccctccatcatgtaccacggc	122						
DB	123	ATCCCTGCCGGGACGAGGAGCCCTGCTCATCGAGGGCGACGCCCTGCTCGATTCTCTACAAGC	182						
QY	123	taactttgacacaaactttgaggacaggaatgatcttgcacggcgcatctgcaaggtacat	182						
DB	183	AAACTTCGATACGAATTTTCGAGGATCGCAATGGATTGTTCACGGGTATCGCCAAGTACAT	242						
QY	183	tgagcaggctacagtcaccctccagcatgaatgagatgctgtggaggaagacatgagtatgc	242						
DB	243	CGAGGAGGGCCACCCACATGCCAACCTTGAATGTCTCTCTGGACGAGGGGGAGAAACACGC	302						
QY	243	ggtcatgctgtacacctggcgagctgttcccgggccattccccaggtgaaatgcaacga	302						
DB	303	AGTAATGCTCTACACCTGGGGTGCTGCTCGCGGCCATTCCGCGAGCCCAAGTCCCAATGA	362						
QY	303	gaagcccaaccgagtagatctatgagaagacagtagaggtgctggcgacggaggtcac	362						
DB	363	GCAGCCGAATCGCGTGAGATCTTACGAGAAGCGGTCGAGGTGCTAGCCCGGAGGTGAA	422						

Query Match 38.7%; Score 346; DB 69; Length 700;
 Best Local Similarity 70.0%; Pred. No. 2.7e-81;
 Matches 463; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY	2	ccatgacgcagcagctccacccctggaaagatgcccgtgccaaacgtggacatgcttgaagagc	61
Db	10	CATGACTACCAACGTCCTCCCTTGAGGATGCCCTTGCCAAATGTGGACTTGTGGAGGACC	69
QY	62	tccccctccccaccagcagccatgcacgtccacgtccaccccttcctccatcatgtaccagg	121
Db	70	TGCCTCTTCAGATCAGCAGCCATGCATGTGAGCAATGCCCTTCACCCCTGCTTATCAG	129
QY	122	ctaaactttgacacaaactttgaggacaggaatgcaattgtcacggcattgcaaggatca	181
Db	130	CGAATATCATCAAAATTTTGAGGACAGAAATGCAATGCTTCACAGGTCTTCCAAAATACA	189
QY	182	ttgagcaggctacagtccactccagcagaatgagatgctggaggaagacatgagtatg	241
Db	190	TTGAGGAGCAACAGTTCATCTGACATGAATGGATGCTTTGAAGAAGAGAAACGCTATG	249
QY	242	cggctcatgctgtacacctggcgagctgttccccgggccattccccaggtgaaatgcaacg	301
Db	250	CGGTTATGCTTACACATGGGTAGCTGCTTCGCGTGTATTCGCAAGTTAAGTGTAAACG	309
QY	302	agcagcccaaccgagtagagatctatgagaagacagtagagggtgctggagccggaggtca	361
Db	310	RACAGTCAACAGGGTTGAGATATACGAAAAACGGTCAAAGTACTTTGAGCCAGAAAGTTA	369
QY	362	caagctcatgaagtctcatgtattttccagcgcaaggccatcgagcggttctcgacgagg	421
Db	370	AGAAGCTGATGGATTTCATGCTCTTTTCAGAAAAAGCGGTGGAAAAAATTTTGGCGTGAAG	429
QY	422	tgaagcggcttgccatgcgcagcgagcgaggagactttgtctctgaggccctactcctga	481
Db	430	TCAGAGAGACTGTGTCAACAGAGAGAGAGAAAGATTTTGNITCCGAAGCTTATCTGTAA	489
QY	482	cccttggcaagtctcatcaacatgtttctgctctgagtagcagctaaagaacatgaagtga	541
Db	490	CTCTCGCAAAATTCATAACATGTTTCGGTATATAGATGACCTGAAAAACATGAAAAAGA	549
QY	542	gcgtcaagaatgaccactccgcctacaagggcgacacagttctctgcggaagatggcag	601
Db	550	GCSTGAAAAACGATTAATTCGGCTTACCGGTAGAGCTGCACAGNTTTTGAAAAAATGCAG	609
QY	602	atccccagctctccaggagctgcagaacctttccatgtctctggtggccaccacaacagga	661
Db	610	ACCGCAGTCCATACAGAATCTCAAAATCTTTTCCATGNTCCTCGCACATCATGACAAA	669
QY	562	t 662	
Db	670	T 670	
RESULT	7		
LOCUS	AI546369	665 bp	mrna
DEFINITION	LD47929.5prime LD Drosophila melanogaster embryo po12 Drosophila melanogaster cDNA clone LD47929 5prime, mRNA sequence.	EST	22-MAR-1999
ACCESSION	AI546369		
VERSION	AI546369.1	GI:4463742	
KEYWORDS	EST.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Drosophila melanogaster		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
TITLE	1 (bases 1 to 665)		
JOURNAL	Harvey,D., Hong,J., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.		
	BDGP/HMI Drosophila EST Project		
	Unpublished (1997)		

```

On May 18, 1998 this sequence version replaced gi:3138128.
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
Plate: 479 row: C column: 5
High quality sequence stop: 501.
Location/Qualifiers
1..565
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD47929"
/clone_lib="LD Drosophila melanogaster embryo pot2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pot2; Site.1: EcoRI; Site.2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. "
BASE COUNT      163 a      175 c      190 g      137 t
ORIGIN
Query Match      38.3%; Score 341.6; DB 47; Length 665;
Best Local Similarity 75.9%; Pred No. 3.9e-80;
Matches 422; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

```

RESULT 8
AU067208
LOCUS

02-JUN-1999

```

DEFINITION AU067208 Sugano mouse brain mncb Mus musculus cDNA clone MNCb-3083
5', mRNA sequence.
ACCESSION AU067208
VERSION AU067208.1 GI:4967945
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 884)
AUTHORS Hashimoto,K., Kusuda,J., Toyoda,A., Tanuma,R., Ito,A., Hirata,M.,
Suzuki,Y., Sasaki,M. and Sugano,S.
TITLE Isolation of full-length cDNA clones from a mouse brain cDNA
library made by oligo-capping method
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189709.
Contact: Katsuyuki Hashimoto
Division of Genetic Resources
National Institute of Infectious Diseases
23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
Email: khashi@nih.go.jp
URL: http://www.nih.go.jp/yoken/genbank/.
FEATURES
source
1..884
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCb-3083"
/lab_host="Sugano mouse brain mncb"
/sex="female"
/dev_stage="adult"
/lab_host="TOP10"
/notes="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
was primed with an oligo(dT) primer-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3.
XhoI sites just outside the DraIII sites can be used to
isolate the cDNA insert. Size selection was performed to
exclude fragments <1.5 kb. Library was constructed by
Sugano et al. (University of Tokyo, Institute of Medical
Science). Custom primer for sequencing: 5' end primer
[CTTCTGCTCTAAAGCTGCG]"
BASE COUNT 241 a 215 c 222 g 193 t 13 others
ORIGIN
Query Match 37.9%; Score 338.6; DB 50; Length 884;
Best Local Similarity 78.9%; Pred. NO. 2.6e-79;
Matches 441; Conservative 0; Mismatches 114; Indels 4; Gaps 3;
QY 4 atgatacgcagctcaccctggaagatgccctgtccacagtgagacctgttgaaagctt 63
DB 88 ATGGGTGCCAAGTACACTAGAGACGCACGTGCCAATGTGGAGCTCTCTGGAGAGCTG 147
QY 64 cccctcccgaccagcagcagctatcagctccacctctccctcatcatgtaccagct 123
DB 148 CTTCTGCTGACACAGCCCTGTGATCGAGGCCCCACCGTCTCTGCTGCTATACGCA 207
QY 124 aactttgcacacaaatttgagacaggaatgcatttgcacggggaattgcaaggtacatt 183
DB 208 AATTTCACACCAACTTTGAACACAGAAATGCAATTTGTACATGGCATTCGAAGATACATC 267
QY 184 gaggcggctacagctccactccagcatgaatgagatgctggaggaaggacatgagtgcg 243
DB 268 GAACAAAGCAACTGCTCCACTCTAGCATGACGAGATGCTAGAGGAAGGCCAAGAGTATGCT 327
QY 244 gtcatgctgtacacgtgcgagctgttcccgccattcccccagtgaaatgcaacgag 303
DB 328 GTCATGCTGTATACCTTGGAGGAGCTGTCTCCCGGGCCATCCCTCAGTGAAGTGCAACGAG 387
QY 304 cagcccaacccagtagatctcatgagaagacagtagaggtgctggagccgaggtcacc 363

```

```

Db 388 CAGCCGAATAGAGTTTGAATTTATGAGAAAAACCGTGGAAAGTCCTTGAACCCGAGGTCACA 447
QY 364 aagctcatgaagttcatgtattttcagcgcagagccatcagcggttctgcagcg-aggt 422
Db 448 AAACGTAGAAATTTATGACTTTTCAGAGAAATGCCATCGAGCGGTCTGTGGGAAAGT 507
QY 423 gaagcggctgtgccatgcgcagcgcaggaaggaacttctctctgagggcctacctctcagc 482
Db 508 GAGACGCCCTGTGTCATGCTGAGAGGAGAAAGGACTTTGTTCTTGAAGCCTCTTGATCAC 567
QY 483 ccttggaaggtcacaacatgttctgtctcctcgtgatgagctaaagaaca-tgaagtga 541
Db 568 CTTGGGCAAAATTT--ATAACATATTTGCTGGGGTTGGATGAACCTCAAGAACATTTGAAGTGA 625
QY 542 gcgtcaagaatgaccactc 560
Db 626 ATGTGAGAGAATGACCACTC 644
RESULT 9
LOCUS AI331713 488 bp mRNA EST 28-DEC-1998
DEFINITION fboic05.y1 zebrafish fin day3 regeneration Danio rerio cDNA 5'
similar to TR:Q14467 Q14467 KIAA0068 ;, mRNA sequence.
ACCESSION AI331713
VERSION AI331713.1 GI:4068272
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 488)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
WashU Zebrafish EST Project 1998
WashU Zebrafish EST Project 1998
Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1796924.
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourcenzentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzp.de)
Possible reversed clone: similarity on wrong strand
Seq primer: T3 Et from Amersham
High quality sequence stop: 411.
Location/Qualifiers
1..488
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish fin day3 regeneration"
/sex="mixed male and female"
/tissue_type="3 day fin regenerates"
/lab_host="E. coli XL0LR"
/notes="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; 1st
strand cDNA primed with (GA)10ACTAGTCGAG(T)18, followed
by second strand synthesis, and ligated to 5' adapter
(5')-aattcgacagag-3', 3'-gccgtgtc-5'. cDNA was cloned
directionally (EcoRI/XhoI) into Stratagene zap express
lambda phage arms. Mass in vivo excision done to obtain

```


REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3188117. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1530 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio1.llnl.gov/bbrp/image/image.html

Insert Length: 876 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 459.

```
FEATURES
  source
    1. 506
      Location/Qualifiers
        seqname="chr1"
        start="1000000"
        end="1000000"
        strand="+"
        type="feature"
        features={
          db_xref="taxon:9606"
          clone="IMAGE:2137854"
          clone_lib="NCI_CGAP_kid"
          db_host="DH10B"
        }
      /organism="Homo sapiens"
    }
  }
}
```

FACSIMILE TRANSMISSION			
BASE COUNT	123 a	156 c	117 g 110 t
ORIGIN			

Query Match	27.6%	Score 246.4;	DB 47;	Length 506;
Best Local Similarity	86.2%;	Pred. No. 5.8e-55;		
Matches 305; Conservative	0;	Mismatches 1;	Indels 48;	Gaps 1;

Qy	1	gccatgatacgcagcgctcaacctgaacgtccctgctccaacgtggaccctcttgaagag	60
Db	153	gccatgaccagcgacgtcacccctggagatgccctgtccaaactggagcctgcttggaagag	212
Qy	61	cttcocctccccgaccagcagccatgatcagcgctccaccttccctcatcatgtaccag	120
Db	213	cttcccctccccgaccagcagccatgatcagcgctccaccttccctcatcatgtaccag	272
Qy	121	gctaactttgacacaacatttgaggacagggaatgcatttgtcacggcgattgcgaagtac	180
Db	273	gctaacctttgacacaaaacttttgaggacagggaatgcatttgtcacggcgattgcgaagtac	332
Qy	181	attgagcaggtcacagtccactccagat-----	210
Db	333	attgacgaggcttacagtcacatccacgtatgatatTTTAGATGCATTGCTCATTTGGCTCC	392
Qy	210	-----gaaatgagatgctggagggaagacatgagtatgcgggtcatgctg	252
Db	393	TCACATCAACCCGTGTGAGAATCAGATGCTGGAGGAAGCACATGATGATGCCGTCTACTGTG	452
Qy	253	tacacctggcgagctgttccocgggccattccccagggtgaaatgcaacagcagc	306
Db	453	TACACCTGGCGCAGCTGTGTCCGGGCCATTCCCAGGTTGAATGCAACGACGAGCAG	506

RESULT	12
AA338114	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)

AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
---------	-------	---------	---------	---------

FEATURES
SOURCE

BASE COUNT	102	a	80	c	99	g	81	t	4	others
ORIGIN										
/organism="Homo sapiens"										
/db_xref="ATCC (inhost):139886"										
/db_xref="taxon:9606"										
/clone_lib="Endometrial tumor"										
/sex="female"										
/dev_stage="adult"										
/note="organ: endometrium; vector: pBluescript SK-;										
Site_1: EcoRI; Site_2: XhoI"										


```
/sex="mixed"
/dev_stage="10-day embryo"
BASE COUNT      89 a      69 c      80 g      82 t
ORIGIN

Query Match      25.3%; Score 225.6; DB 59; Length 320;
Best Local Similarity 81.6%; Pred. No. 1.6e-49;
Matches 261; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 387 tcagcgcaaggccatcgagcgttcctgcgcgaggtgaagcgctgtgcacatgccagcg 446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TCAGAGAAATGCGATCGAGCGGTCTCTGTGGGAACGTGACACGCTGTGTCTGTGAGAG 60

QY 447 caggaaggacttctctgagggcctaccctcctgacccttgccactgtgcacatggtt 506
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GAGAAAGGACTTTGTTCTGAAGCCCTACTGTGATCACCCCTGGGGGAAATTTATCAACATGTT 120

QY 507 tgcgtcctggatgagctaaagacatgaagtgcagcgtcaagatgaccactccgccta 566
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TGTGTGTTGGATGAGCTGAAGAACATGAAGTGCATGTGGAAGATGACCACCTCTGCATA 180

QY 567 caagggcgagcacagtctcctgcgggaagtggcagatccccagctctatccaggagtcgca 626
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TAAGAGGGCTGCTCAGTCTTTTACGTAAATGGCAGATCCACAAATCCATCCAGGAGTCACA 240

QY 627 gaaccttccatgttctcggcccaaccacacagcagatcaccacagctgtctccaccagcaact 686
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GAATCTGTCCATGTTCTTGCCCAACCAACACAAAGATTACAAATCTCTGCAGCAGCAGCT 300

QY 687 tgaagtgtatccaggctatg 706
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TGAGGTCAITTCGGGTATG 320

Search completed: May 23, 2000, 22:03:32
Job time: 3970 sec
```

```
RESULT 15
AI834230/c
LOCUS      AI834230      217 bp      mRNA      EST      13-JUL-1999
DEFINITION RCO-HT0012-030799-001-H04 HT0012 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AI834230
VERSION     AI834230.1 GI:5456573
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  HCGP http://www.ludwig.org.br/ORESTES.
            The FAPESP/LICR Human Cancer Genome Project
            Unpublished (1999)
COMMENT    On May 18, 1998 this sequence version replaced gi:3136806.
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RCO&t2=RCO-HT0012-
            030799-001-H04&t3=1999-07-03&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 47
            High quality sequence stop: 217.
            Location/Qualifiers
                1..217
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="HT0012"
                    /dev_stage="adult"
                    /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:03:32 ; Search time 3267.04 Seconds
(without alignments)
558.288 Million cell updates/sec

Title: US-09-215-435-116
Perfect score: 450
Sequence: 1 ctgtctccagcgtgacgcg.....acggacaagtcttagacag 450

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
1	450	100.0	476	33	AA452724	AA452724 zx39d11.r
2	438	97.3	573	36	AA622037	AA622037 nq23b11.s
3	428	95.1	591	79	AL135416	AL135416 DRFZP762L
4	421.8	93.7	509	44	AI264864	AI264864 qd89a05.x
5	414	92.0	500	40	AA931276	AA931276 q851a01.s
6	411	91.3	489	43	AI200921	AI200921 gf63b11.x
7	410.8	91.3	498	42	AI125964	AI125964 qc46b07.x
8	409.2	90.9	485	44	AI263705	AI263705 qg90b11.x
9	408	90.7	408	31	AA314244	AA314244 EST186157
10	407.8	90.6	500	43	AI187865	AI187865 qc08g02.x
11	397	88.2	468	36	AA643309	AA643309 nr59b05.s
12	392.6	87.2	548	29	AA156940	AA156940 z119d01.s
13	392	87.1	392	33	AA416757	AA416757 zu09a01.s
14	390.2	86.7	468	44	AI264865	AI264865 qd89a06.x
15	388.8	86.4	514	39	AA903588	AA903588 qk51g01.s
16	363.4	80.8	373	40	AA948665	AA948665 qd4a03.s
17	362.4	80.5	462	29	AA133412	AA133412 zx96e04.r
18	356.8	79.3	420	25	W07240	W07240 za94f08.r1
19	354.4	78.8	527	28	AA115385	AA115385 zk96e04.s
20	350	77.8	423	35	AA557190	AA557190 n174g09.s
21	345	76.7	510	39	AA903051	AA903051 qk51e01.s
22	327.8	72.8	454	39	AA870917	AA870917 vq25a06.r
23	325.8	72.4	411	39	AA870712	AA870712 vq25a06.r
24	322	71.6	426	22	R81019	R81019 y194g04.r1
25	320.2	71.2	445	28	AA089135	AA089135 mo21b03.r
26	318.2	70.7	438	34	AA500885	AA500885 vg03b01.r
27	317.2	70.5	442	30	AA253672	AA253672 mx84e03.r
28	316.8	70.4	342	22	R71425	R71425 y151c06.r1
29	315.8	70.2	427	22	R80662	R80662 y193e05.s1
30	309	68.7	551	30	AA241291	AA241291 mv31g07.r
31	308	68.4	448	36	AA611548	AA611548 vo85b12.r
32	307.6	68.4	444	39	AA887693	AA887693 nr10f05.s
33	306	68.0	645	106	AQ391270	AQ391270 CIRBI-E1-
34	302.8	67.3	383	27	AA027638	AA027638 mi09e02.r
35	302	67.1	429	107	AQ453378	AQ453378 HS_5192_A
36	300.8	66.8	456	27	AA043775	AA043775 zk31h11.s
37	298.2	66.3	367	39	AA903227	AA903227 ok34d01.s
38	293.4	65.2	475	69	AW113705	AW113705 MC1938 mo
39	283.8	63.1	288	22	R80461	R80461 y193e05.r1
40	282.6	62.8	419	30	AA208293	AA208293 mv96g02.r
41	281.8	62.6	441	36	AA608360	AA608360 vm89h01.r
42	278.8	62.0	345	43	AI1172308	AI1172308 EST218309
43	275.4	61.2	393	27	AA032038	AA032038 zk18a07.s
44	274.6	61.0	433	27	AA003283	AA003283 mg42c03.r
45	271.4	60.3	358	36	AA639204	AA639204 ns04b05.r

ALIGNMENTS

RESULT 1
AA452724 476 bp mRNA EST 05-JUN-1997
LOCUS AA452724.1 Scores_total_fetus_Nb2HF8_9w Homo sapiens cdna clone
DEFINITION IMAGE:788853 3', mRNA sequence.
ACCESSION AA452724
VERSION AA452724.1 GI:2166393
KEYWORDS BSL
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 476)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
Moore, B., Scheinberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wyllie, T., Waterston, R. and Wilson, R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402122.
Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 457.
Location/Qualifiers
1. .476
/organism="Homo sapiens"
/db_xref="GDB:5986470"
/db_xref="taxon:9606"
/clone="IMAGE:788853"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 175 a 93 c 124 g 84 t
ORIGIN

Query Match 100.0% Score 450; DB 33; Length 476;
Best Local Similarity 100.0%; Pred. No. 6.4e-96;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctgtctccagcgtgacgcccagccatgctgacgagagcttgcgctgagagacag 60
Db 6 CTGCTCCAGCGGTGACGCGGACCATGCGGACGAGAGCTTGAGCGCTGAGGAGACG 65
QY 61 aggtcggcagcgtcagcccaaacacacggggtatcttgcgctgagagacagca 120
Db 56 AGGCTGGCGGAGCTGACGCGCAACACACGGGATCTGTGTGATGGCGGCCAACAGCA 125
QY 121 aagcacagggagcagaatgagaaacagtagtatcttgcgctgagagacagcgc 180
Db 126 AAGCACAGGGAAGCAGAAATGAGAAACAGTATCTAGCCCAAGTCTGGATCATCGGCC 185
QY 181 cggccaggttaagtaactagcctgttaagcctgaaagcctgaaagcagtagagaat 240
Db 186 CGGCCAGGTTAAGTAACTTACCTTAAAGCCTGAAAAAAGCTAAAGCAGTAGAANT 245
QY 241 tacctatcacagatggaagatggaacataagtagagaggttatcagaaagggtta 300
Db 246 TACCTTATACAGATGGCAAGATATGGCAACTAAGTAGAAGGTATCAGAACAGGTTA 305
QY 301 atagaaatctcttaaaaaagtaagcccaacacagaaagacacacagtgaaattcaac 360
Db 306 ATAGAAATCTCTTAAAAAAGTAAGCCCAACAAACAGAAAGACACACAGTGAATTCAC 365
QY 361 agaagaaagtaagtagcctctgtagagagtagagaggttatcagaaagggtta 420
Db 366 AGAAGAAAGTATGGACTCTGATGAAGATGAGGATATTGAACTACAACTGCTCAGAGA 425
QY 421 ctgaacttaacggaacaaagctctaggacag 450
Db 426 CTAGAACTTAACGGAAACAAGCTTAGGACAG 455

RESULT 2
AA622037/c 573 bp mRNA EST 31-OCT-1997
LOCUS AA622037.1 NCI_CGAP_Col0 Homo sapiens cdna clone IMAGE:1144701 3',
DEFINITION nq23b11.s1

mRNA sequence.
 AA622037
 AA622037.1 GI:2525913
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 573)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Sep 12, 1996 this sequence version replaced gi:1404975.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 592 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 446.
 Location/Qualifiers
 1..573
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1144701"
 /clone_lib="NCI-CGAP_Col0"
 /tissue_type="colon tumor RER+"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER+ colon tumor, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo (N-Soares4)."
 BASE COUNT 109 a 143 c 107 g 214 t
 ORIGIN
 Query Match 97.3%; Score 438; DB 36; Length 573;
 Best Local Similarity 100.0%; Pred. No. 4.3e-93;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 tgacgcgagccatgacgagagagcttgagcgctgagagacagagctggccgag 72
 DB 510 TGACGCGAGCCATGCGGAGGAGGCTTGAGCGCTGAGGACAGAGCGTGGCCGAG 451
 QY 73 ctgcaggccaaacacggggatctctgtgctgcgcccacacaggaaagacacaggaa 132
 DB 450 CTGACGCGCCAAACACGGGGATCTCTGTGTATCGCGCCCAACAGGAAGCAACAGCAGGGAA 391
 QY 133 qcagaaatgaaacagatcttagcccaagtctgcatcagtcgcccggccagatta 192
 DB 390 GCAGAAATGAGAAACAGATCTTAGCCCAAGTCTTGATCAGTCGGCCCGCCGAGTTA 331
 QY 193 agtaacttagcaactgttaaagcctgaaaaaataaagcagtagagaattaccttatacac 252
 DB 330 AGTAACCTAGCACTGTAAAGCCTGAAAAAATAACAGTAGAGATTACCTTATACAG 271
 QY 253 atggcaagatatggacaactaagttagaaggtatcagaacaaggtttaaataagaaatcctt 312
 DB 270 ATGCAAGATATGGACAACCTAAGTGAGAAGTATCAGAACAAAGGTTTAAAGAAATCCTT 211

QY 313 aaaaagttaagccaacaaacagaaaaacacacacagtgaaattcaacagagaagaaagta 372
 DB 210 AAAAAAGTAAAGCCAAACAAACAGAAAAGACACACACAGTGAATTCACAGAGAAGAAAGTA 151
 QY 373 atggactctgatgaagatgaagattattgaactacaagtgtcactacagactagaacttaac 432
 DB 150 ATGACGCTCTGATGAAGATGACGATTATTGAAGTCAAGTGTCTACAGACTAGAACTTAAC 91
 QY 433 ggaacaagtctaggacag 450
 DB 90 GGAACAAGTCTAGGACAG 73
 RESULT 3
 AL135416
 LOCUS AL135416 591 bp mRNA EST 20-DEC-1999
 DEFINITION DKFZP762L1315_r1 762 (synonym: hmel2) Homo sapiens cDNA clone
 DKFZP762L1315 5', mRNA sequence.
 ACCESSION AL135416
 VERSION AL135416.1 GI:6603603
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 591)
 AUTHORS Ansorge W., Wirkner U., Mewes W., Weill B. and Wiemann S.
 TITLE EST (Ansorge W., Wirkner U., Mewes W., Weill B. and Wiemann S.)
 JOURNAL Unpublished (1999)
 COMMENT On Jul 7, 1999 this sequence version replaced gi:5408221.
 Contact: Ansorge W
 MIPS
 Am Klopferspitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by EMBL (European Molecular Biology Laboratories,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 No sl sequence available.
 This clone is available at the RZPD in Berlin..
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 FEATURES
 Location/Qualifiers
 1..591
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZP762L1315"
 /clone_lib="762 (synonym: hmel2)"
 /tissue_type="melanoma (Mewo cell line)"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
 BASE COUNT 223 a 109 c 146 g 113 t
 ORIGIN
 Query Match 95.1%; Score 428; DB 79; Length 591;
 Best Local Similarity 99.6%; Pred. No. 9.6e-91;
 Matches 450; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
 QY 1 ctgctccagcgctgaacccgagccatggcgagcagagagcttgaggcgctgaggagacag 60
 DB 40 CTGCTCCAGCGCTGACCCGAGCCGAGCATGGCGGAGGAGCTTGAGGCGCTGAGGAGACAG 99
 QY 61 aggtgcccagctgcagggccaaacacggggatctctggtgatcgcccaacagagaagca 120
 DB 100 AGGTGCGCGAGCTGCAGGCCAAACACGGGGATCTCTGTTGATGCGGCCACACAGGAACA 159
 QY 121 aagcacaggg-aagcagaaatgagaacagtagtatcttagcccaag-ttctggatcagtcgg 178
 DB 160 AAGCACAGGGCAAGCAGAAATGAGAAACAGTAGTATCTTAGCCCCAAGTTTCTGGATCAGTCGG 219

QY	179	ccggggccaggttaagtaacttagcactgtgtaagcctgaaaaaactaaagcagtagaga	238
Db	220	CCGGGGCCAGGTTAAAGTAACTTAGCAGCTTCTAAAGCCTGAAAAAACTAAAGCAGTAGAGA	279
QY	239	attaccttatacagatggcaagatatatggacaactaagtgaaggtatcagaacaaggtt	298
Db	280	ATTACCTTTATACAGATGGCAAGATATGGCACTTAAGTGAGAGGTTATCAGAACAAGGTT	339
QY	299	taatagaaatccttaaaaaagtaagcccaacaacagaaaaagacaacaacacagtgaaaaatca	358
Db	340	TAATAGAAATCCTTAAAAAAGTAAGCCAAACAACAGAAAAAGACAACAACAGTGAATTC	399
QY	359	acgaagaagaagtaatggactctgatgaagatgacgattattgaactacaagtgctcaaca	418
Db	400	ACAGAGAAAAAGTAATGGACTCTGATGAAGATCAGGATTTGAAGTACAAAGTGCTCACA	459
QY	419	gactgaacttaacgaacaagctaggacag	450
Db	460	GACTAGAACTTAACGGAAACAGTCTAGGACAG	491
RESULT	4		
LOCUS	AI264864/C		
DEFINITION	AI264864 509 bp mRNA EST 13-NOV-1998 Q989A05.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:1938512 3' similar to TR:O14737 O14737 TFAR19.1; mRNA sequence.		
ACCESSION	AI264864.1 GI:3873067		
VERSION	AI264864		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 509) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2152815. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq Primer: -400P from Gibco High quality sequence stop: 408. Location/Qualifiers		
FEATURES	source 1..509 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1938512" /dev_stage="8-9 weeks" /lab_host="DH108" /note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGGCGCTAAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "		
BASE COUNT	104 a	116 c	88 g
ORIGIN	200 t	1 others	
Query Match	93.7%	Score 421.8;	DB 44; Length 509;
Best Local Similarity	99.3%	Pred. No. 2.7e-89;	
Matches 423; Conservative	0; Mismatches 3;	Indels 0; Gaps 0;	

BASE COUNT	104 a	118 c	86 g	192 t
ORIGIN	/lab_host="DH10B" /note=vector: p7T73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T73 vector. This library is the normalized version of NCI-CGAP-Brl.1. Library was constructed by Bento Soares and M. Fatima Bonaldo."			

Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbbr/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 409.
Location/Qualifiers
1. .489

```

FEATURES
  source
    1. 489
    Location/Qualifiers
      db_xref="taxon:9606"
      clone_lib="IMAGE:1754685"
      clone_lib="Soares_testis_NHT"
      sex="male"
      lab_host="DH10B"
      note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTTACCAATCTCAAGTCGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Col5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
103 a 111 c 84 g 191 t

```

Query Match	91.3%	Score 411;	DB 43;	Length 489;
Best Local Similarity	100.0%;	Pred. No. 9.1e-87;		
Matches 411;	Conservative 0;	Mismatches 0;	Indels 0;	Caps 0;

Qy	40	cttgaggcgtgaggaacagagagctggccgagctcagagccaaacacggggatctcgggt	99
Db	489	CTTGAGGCGGTGAGGAGACAGAGGCTGGCCGAGCTGCAGGCCAAACACGGGGATCCTGGGT	430
Qy	100	gatcgggcccaacaggaagcaaaacacagggaaacagaaaatgagaacagtatcttagcc	159
Db	429	GATCGGCGCCCAACAGGAAGCAAAAGCACAGGGAACAAATGAGAAACAGTATCTTAGCC	370
Qy	160	caagttcttgatcagtcggtcgccggccgaggttaagtaacttagcactgttaaaagcctgaa	219
Db	369	CAAGTTCTTGATCAGTCGGCCGGCCGAGGTTAAGTAACTTAGCACTTGTAAGACCTGAA	310
Qy	220	aaaactaaacagtagagaattacccttatcacagatggcgaagatatggacaactaagttag	279
Db	309	AAAACCTAAACAGCTAGAGAATTACCTTTATACAGATGGCAAGATATGGACAACCTAAGTGA	250
Qy	280	aaggtatcagaacaaggtttaaatagaactcttaaaaaagtaagccaaacacagaagaag	339
Db	249	AGGTATCAGAACACANGGTTTTATAGAAATCCTTTAAAAAAGTAAGCCCAACAAACAGAAAG	190
Qy	340	acaacaacagdtgaaattcacagaagaagaaagtatggactctgatgaagtgcagattat	399
Db	189	ACAACAACAGTGAAATTCACAGAGAAGAAAAGTAATGGACTCTGTATGAATGATGACGATTAT	130
Qy	400	tgaactcaagtgctcacagactagaaacttaacggaaacagtgcttagagacag	450
Db	129	TGAACCTACAAGTGCTCACAGACTAGAACTTTAACGGAAACAAGTCTTAGACACAG	79

RESULT	7
LOCUS	AI125964/c
DEFINITION	AI125964 498 bp mRNA EST 27-OCT-1998 qc46b07.x1 Soares.pregnant.uterus.NbHPU Homo sapiens cDNA clone IMAGE:1712629 3' similar to TR:014737 014737 TFR19. ;, mRNA sequence.
ACCESSION	AI125964
VERSION	AI125964.1 GI:3594478

RESULT	6	
AI200921/c		
LOCUS		
DEFINITION	AI200921 489 bp mRNA EST 14-OCT-1998 qf63b11.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1754685 3', similar to TR:O14737 O14737 TFAR19. ;, mRNA sequence.	
ACCESSION	AI200921	
VERSION	AI200921.1	GI:3753527
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 489)	
TITLE	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	On Jan 14, 1998 this sequence version replaced gi:1877663. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima	

[illegible]

```
QY 277 gagaaggtatcagaacaaggttttaataagaaatccttaaaaaagtaagccaaacacagaa 336
|||||
Db 245 GAGAGGTATCAGAACAGAGTTTAAATAGAATCTTAAAAAAGTAAGCCCAACAAACAGAA 186
|||||
QY 337 aagacaacaacagtgaaattcaacagaagaataatggactctgatgaagatgacgat 396
|||||
Db 185 AAGACAACAACAGTGAATTCACAGAGAAGAAAAGTAATGGACTCTGATGAAGATGACGAT 126
|||||
QY 397 tattgaactacaagtgctcacagactagaacttaacgggaacaaagtctaggacag 450
|||||
Db 125 TATTGAACATAAGTGCTCACAGACTAGAACTTAACGGAAACAAGTCTAGGACAG 72

RESULT 9
AA314244
LOCUS AA314244 408 bp mRNA EST 19-APR-1997
DEFINITION EST186157 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5'
end, mRNA sequence.
ACCESSION AA314244
VERSION AA314244.1 GI:1966645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 408)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Melssner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 12140200
COMMENT On May 8, 1995 this sequence version replaced gi:801519.
Other ESTs: THC169754
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1..408
/organism="Homo sapiens"
/db_xref="ATCC (inhost):110913"
/db_xref="taxon:9606"
/clone_lib="Colon carcinoma (HCC) cell line II"
/tissue_type="colon"
/cell_type="KM12C"
/cell_line="KM12C(HCC)-parental human colon
carcinoma;Dukes B2"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
```

```
BASE COUNT 151 a 81 c 108 g 68 t
ORIGIN
Query Match 90.7%; Score 408; DB 31; Length 408;
Best Local Similarity 100.0%; Pred. No. 4.5e-86;
Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctctccagcgtgcagccagcgcagcagcagcagcagcagcagcagcagcagcagcagcag 60
|||||
Db 1 CTGCTCCAGCGCTGACGCCGAGCCATGGCGGAGGAGGAGCTTGAGGCGCTGAGGAGACAG 60
|||||
QY 61 agctggcagcgtgcagcgcagcgcagcgcagcgcagcgcagcgcagcgcagcgcagcagca 120
|||||
Db 61 AGGCTGGCCGAGCTGCAGGCGCCAAACACGGGGATCCTGTGTGATGCGGCCCAACAGGAACA 120
|||||
QY 121 aagcacagggaagcagaaatgagaacacagtatctaggcccaagtctctggatcagtcggcc 180
|||||
Db 121 AAGCACAGGGAAGCAGAAATGAGAAACAGTATCTTAGCCCAAGTCTGTGATCAGTCGCGCC 180
|||||
QY 181 cgggccaggttaagtaacttagcacttgtaaaccttgtaaaccttgtaaaccttgtaaacct 240
|||||
Db 181 CGGGCCAGGTTAAGTAACTTAGCCTTGTAAAGCCTTAAAGCCTTAAAGCAGTAGAGAAT 240
|||||
QY 241 tacctatcacagtgccaagatatggacaactaagtgagaagtgatcagacaaggttta 300
|||||
Db 241 TACCTTATACAGATGGCAAGATATGGCAACTAAGTGAGAGGTATCAGAACAGGTTTA 300
|||||
QY 301 atagaatcctctaaaaaagtaagcccaacacagaaagacacacacacagtgaaattcaac 360
|||||
Db 301 ATAGAAATCCTTAAAAAAGTAAGCCAAACACAGAAAGACAAACACAGTGAAATTCAC 360
|||||
QY 361 agaagaaaaagtaagcactctgatgatgaagcagcagcagcagcagcagcagcagcagcag 408
|||||
Db 361 AGAAGAAAGTAATGGACTCTGATGAAGATGACGATTATTGAACTACA 408
|||||

RESULT 10
AA187865/c
LOCUS AA187865 500 bp mRNA EST 30-OCT-1998
DEFINITION qe08g02.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1738418
3' similar to TR:O14737 O14737 TFAR19. ; mRNA sequence.
ACCESSION AA187865
VERSION AA187865.1 GI:3739074
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 500)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/brrp/image/image.html
Insert Length: 596 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 317.
Location/Qualifiers
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1738418"
```



```
DEFINITION z119d01.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:502369 3', mRNA sequence.
ACCESSION AA156940
VERSION AA156940.1 GI:1728555
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 548)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoso,S., Dietrich,N., Dubuque,T., Faveillo,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On Sep 12, 1996 this sequence version replaced gi:1392764.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 595 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 431.
FEATURES
source
1..548
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:3807357"
/db_xref="taxon:9606"
/clone="IMAGE:502369"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
ACTGCGAGATTCGCGCCGCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
BASE COUNT 108 a 134 c 105 g 196 t
ORIGIN
Query Match 87.2%; Score 392.6; DB 29; Length 548;
Best Local Similarity 96.4%; Pred No. 1.9e-82;
Matches 432; Conservative 0; Mismatches 12; Indels 4; Gaps 3;
QY 5 tcacgcgtgacgccgagcc--atgcgcgacgagcgcttgagcgctgagcgacagag 62
||| |||||
DB 519 TCCACGCTGACGCCGACCCATGCGCNGACCGAGGAGCTTGAGCGCTGAGGAGACAGAG 460
QY 63 gctgcgcgctgcagcgaacacacaggggacgctgctgctgctgcgcgcacagaaacaa 122
||| |||||
DB 459 NCTGGCCGAGCTGC-GGCCAAACACACGGGATCCTGCTGATGCGCCCAACAGGAAGC-AA 402
QY 123 gcaagggaagcagaatacagaaacagctcttagcccaagttctgagtcagtcgcccgc 182
||| |||||
DB 401 GCACGGGAGCAGCAANTGAGAACAGTAGTCTTTAGCCCCAAGTTCTGGATCAGTCGCCCG 342
||| |||||
QY 183 ggcacggttaagttaactgacacttgtaagccttaagccttaagccttaagccttaagcctta 242
||| |||||
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:03:49 ; Search time 3267.04 Seconds
(without alignments)
973.903 Million cell updates/sec

Title: US-09-215-435-118
Perfect score: 785
Sequence: 1 cggaatccggagtcggtg.....tatggaataaagtcttttc 785

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

VERSION AI028601.1 GI:3245910
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 483)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicqap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1796903.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 982 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 414.
 FEATURES
 source Location/Qualifiers
 1..483
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1644098"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5'
 TGTACCAATCTGAAGTGGAGGCGCGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 77 a 143 c 148 g 115 t
 BASE COUNT 77 a 143 c 148 g 115 t
 ORIGIN
 Query Match 50.1%; Score 393; DB 41; Length 483;
 Best Local Similarity 100.0%; Pred. No. 6.4e-78;
 Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 393 aaggtaccagatggaggagaagggccctggtaccatccagaaggccacggacagc 452
 Db 406 AAGGTACCAGGATGGAGGAGGAGGCGCCCTGTACCATCCAGAGGCCACGGACAGC 347
 Qy 453 ttccacagaaactccatccccggggtggccttctggatcattagctgcccagggagg 512
 Db 346 TTCCACACAGAACTCCATCCCGGGGTGGCCCTCTGGATCAATTAAGTGCACACGGCGGAGG 287
 Qy 513 tccacaggatgacctggaggcgccactggctcagcagagaagcacaccgctcgag 572
 Db 286 TCCACACAGGATGCTTGGAGGGGGGGCCACTGGCTCAGCAGAGGACACCCGGCTGCAG 227
 Qy 573 gccatccgggtgactccgcgaaggggaccacacagagcgctctagaagaggagaccgag 632
 Db 226 GCCATCCGGGTGACTCCGCAAGGGGACCCACAAAGGACGTCCTAGAGAGGGACCCGAG 167
 Qy 633 agctctccactcaggctgtccccccgaaagacccacttactgtacatctcaggccc 692
 Db 166 AGCTCTCCACTCAGGCTGTCCCCCGGAAAGACCCACTTACTGTACATCCTCAGGCC 107

Qy 693 tctcggcagctgtagggtgggaccgggagcagcctgttagcccccatcacacct 752
 Db 106 TCTCGCAGCTGTAGGGGTGGGACCGGGAGACCTGCTGTAGCCCATCAGACCT 47
 Qy 753 gccccagcaccatattgaaataaagtctcttc 785
 Db 46 GCCCAAGCACCATATGGAATAAAGTTCTTTC 14
 RESULT 3
 AW162347/c
 LOCUS
 DEFINITION au74c11.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
 IMAGE:2782004 3', similar to SW:LFE4.CHICK Q90839 UNKNOWN LENS FIBER
 PROTEIN CLFEST4 PRECURSOR. ;, mRNA sequence.
 ACCESSION AW162347
 VERSION AW162347.1 GI:6301380
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 598)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maria,M.,
 Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI Human EST Project
 Unpublished (1997)
 On Jul 7, 1999 this sequence version replaced gi:5407190.
 Other ESTs: au74c11.y1
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Glbco
 High quality sequence stop: 464.
 Location/Qualifiers
 1..598
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2782004"
 /clone_lib="Schneider fetal brain 00004"
 /sex="male"
 /tissue_type="frontal lobe"
 /dev_stage="5 months post-conception"
 /lab_host="DH10B"
 /note="organ: brain; Vector: pBluescript SK (Stratagene);
 Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
 prepared from human fetal brain tissue. 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence:
 5'-GAGAGAGAGAGAGACTCAAGATCTTTAATTAATTAATCAATCCCCCCCCCCC-3'
 and 3' adaptor sequence:
 5'-GAGAGAGAGACTCGAGTGTGTGTGTGTGTGTGTGTGT-3'. The library was
 size-selected for >0.5 kb inserts and has an average
 insert size estimated at 1.2 kb. This library was
 constructed using the CAP-trapper method for full-length
 enrichment and has not undergone amplification. Library
 was constructed by Dr. Claudio Schneider (LNCIB-Area
 Science Park, Trieste, Italy)."
 91 a 168 c 184 g 155 t
 BASE COUNT 91 a 168 c 184 g 155 t
 ORIGIN
 Query Match 50.1%; Score 393; DB 72; Length 598;
 Best Local Similarity 100.0%; Pred. No. 6.9e-78;
 Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 393 aaggtaccaggatgaggagaaggagccctgttaccatccagaagggccacggacgc 452
Db 409 AAGTACCCAGGATGGAGGAGAGAGGCGCCCTGGTACCCATCCAGAAGCCACGACAC 350
QY 453 ttccacagaactccatcccccgggtggcctctgagatcaataagctgcacggcgagg 512
Db 349 TTCCACACAGAACTCCATCCCGGGTGGCCTTCTGGATCATTAAGCTGCACGGCGAGG 290
QY 513 ttccacacagatccctgagggcgccactggctcagcgagaagcgacacgcgcgcag 572
Db 289 TCCACACAGGATCCCTGGAGGGCGGCCACTGGCTCAGCGAGAGCGACACCGCTGCAG 230
QY 573 gccatccggatggactccgcgaagggaacccacaaggagcgtcctaagaggggacccgcag 632
Db 229 GCCATCCGGATGGACTCCGCAAGGGAGCCCAAGAGCGTCTAGAAAGAGGGGACCGAG 170
QY 633 agctctccactcaggctgtccccccgaagacccacttactctacatcctcaggccc 692
Db 169 AGCTCTCCCACTCCAGGCTGTCCCCCGAAGAGCCCACTTACTGTATCTCCTCAGGGCCC 110
QY 693 tctcgacagttaggggtgggacccggggacacctgctgtgagcccccatcagacct 752
Db 109 TCTGGCAGCTGTAGGGGTGGGGACCGGGAGCACCTGCCTGTAGCCCCCATCAGACCT 50
QY 753 gcccacagaccatgaggaataaagtcttctc 785
Db 49 GCCCCAAGCACCATATGAATAAGTTCTTTC 17

RESULT 4
A1150592/c
LOCUS
DEFINITION qf36g11.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1752164
3', mRNA sequence.
ACCESSION A1150592
VERSION A1150592.1 GI:3679061
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 428)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397845.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldio, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 506 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 419.
Location/Qualifiers
1..428
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1752164"
/clone.lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
note=Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)

```

```

primer [5'
TGTTACCAATCTGAAGTGGAGGGCGGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldio. "
BASE COUNT 67 a 126 c 135 g 99 t 1 others
ORIGIN

Query Match 49.9%; Score 392; DB 42; Length 428;
Best Local Similarity 99.7%; Pred. No. le-77;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 393 aaggtaccaggatgaggagaaggagccctgttaccatccagaagggccacggacgc 452
Db 404 AAGGTACNAGGATGGAGGAGAGAGGCGCCCTGGTACCCATCCAGAAGCCACGACAGC 345
QY 453 ttccacagaactccatcccccgggtggcctctgagatcaataagctgcacggcgagg 512
Db 344 TTCACACAGAACTCCATCCCGGGTGGCCTTCTGGATCATTAAGCTGCACGGCGAGG 285
QY 513 tccacacagatgcctggagggcgccactggctcagcgagaagcgacacgcgcgcag 572
Db 284 TCCACACAGGATGCCCTGGAGGGCGGCCACTGGCTCAGCGAGAGCGACACCGCCCTGCAG 225
QY 573 gccatccggatggactccgcgaagggaacccacaaggagcgtcctaagaggggacccgcag 632
Db 224 GCCATCCGGATGGACTCCGCAAGGGAGCCCAAGAGCGTCTAGAAAGGGGACCGAG 165
QY 633 agctctccactcaggctgtccccccgaagacccacttactctacatcctcaggccc 692
Db 164 AGCTCTCCCACTCCAGGCTGTCCCCCGAAGAGCCCACTTACTGTATCTCCTCAGGGCCC 105
QY 693 tctcgacagttaggggtgggacccggggacacctgctgtgagcccccatcagacct 752
Db 104 TCTGGCAGCTGTAGGGGTGGGGACCGGGAGCACCTGCCTGTAGCCCCCATCAGACCT 45
QY 753 gcccacagaccatgaggaataaagtcttctc 785
Db 44 GCCCCAAGCACCATATGAATAAGTTCTTTC 12

RESULT 5
A1879701/c
LOCUS
DEFINITION A1879701 460 bp mRNA EST 23-AUG-1999
IMAGE:2518866 3', mRNA sequence.
ACCESSION A1879701
VERSION A1879701.1 GI:5553750
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 460)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Krizman, D., Kubacka, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On Feb 18, 1999 this sequence version replaced gi:4296363.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco.

```


FEATURES
 source Location/Qualifiers
 1. .460
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2518866"
 /clone_lib="Schneider fetal brain 00004"
 /sex="male"
 /tissue_type="frontal lobe"
 /dev_stage="5 months post-conception"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript SK (stratagene);
 Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
 prepared from human fetal brain tissue. 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence:
 5'-GAGAGAGGAAGAGCTCAAGAGTCCTTAATTAATTAATCCCGCCGCCGCC-3'
 and 3' adaptor sequence:
 5'-GAGAGAGAGACTCGAGTTTCTTTTCTTTT-3'. The library was
 size-selected for >0.5 kb inserts and has an average
 insert size estimated at 1.2 kb. This library was
 constructed using the CAP-trapper method for full-length
 enrichment and has not undergone amplification. Library
 was constructed by Dr. Claudio Schneider (LNCIB-Area
 Science Park, Trieste, Italy)."
 BASE COUNT 73 a 133 c 142 g 111 t 1 others
 ORIGIN
 Query Match 49.9%; Score 392; DB 62; Length 460;
 Best Local Similarity 99.7%; Pred. No. 1.1e-77;
 Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 393 aaggtaccaggatggaggagaagggccctggtaccatccagaagggccacggacagc 452
 |||||
 DB 409 AAGGTACCCAGGATGGAGGAGAAGAGGCCNTGTTACCATCCAGAGGCCACGGACAGC 350
 |||||
 QY 453 ttccacacagaatccatccccgggtgctcttggtatcataagctgccacgcygaggg 512
 |||||
 DB 349 TTCCACACAGAACTCCATCCCGGTGGCTTCTGTGATCATTAAGCTGCCACGGCGGAGG 290
 |||||
 QY 513 tccacacaggatgccctggaggggcgccactggctcagcagaagacacggcctgcag 572
 |||||
 DB 289 TCCACACAGGATGCCCTGGAGGGCGGCCACTTGGCTCAGCGAGAAGCACACCGCTGCAG 230
 |||||
 QY 573 gccatccggatggactccgcgaaggggaccacaaagacgctcctagaaggggacccag 632
 |||||
 DB 229 GCCATCCGGGTGACATCCCGCAAGGGAGCCCAAGAGACGTCCTTAGAAGGGGACCGAG 170
 |||||
 QY 633 agctctctccactccaggctgtccccccgaagaccacttactgtacatctcagggccc 692
 |||||
 DB 169 AGCTCTCCCACTCCAGGCTGTCCCGCCGAAAGACCCACTTACTGTACATCTCTCAGGCC 110
 |||||
 QY 693 tctcggcagctgtagggggtggggaccggggagacacctgcctgtagggcccaatagaccct 752
 |||||
 DB 109 TCTCGCGACGCTGTAGGGGTGGGGACCGGGGAGCACCTGCTGTAGCCGCCCATCAGACCT 50
 |||||
 QY 753 gccccaaagcaccataggaaataaagttctttc 785
 |||||
 DB 49 GCCCCAGACCAATATGGAAATAAGTCTTTTC 17
 |||||
 RESULT 6
 AI200868/c
 LOCUS
 DEFINITION AI200868 428 bp mRNA EST 14-OCT-1998
 qf62g12.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1754662
 3', mRNA sequence.
 ACCESSION AI200868
 VERSION AI200868.1 GI:3753474
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 428)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jan 14, 1998 this sequence version replaced gi:1877579.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 406.
 Location/Qualifiers
 1..428
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:1754662"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5',
 TGTATCAACATCTGAAGTGGAGCGCGCGCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 67 a 125 c 136 g 100 t

BASE COUNT
 ORIGIN

Query Match	49.9%	Score	391.4;	DB	43;	Length	428;
Best Local Similarity	99.7%;	pred. No.	1.4e-77;				
Matches	392;	Conservative	0;	Mismatches	1;	Indels	0;
Gaps	0;						

```

Qy 393 aaggtaccaggatgaggagagaagagggccctgtaccaccaccgaagggccacggacagc 452
      |||
Db 404 AAGGTACCAGGATGGAGGAGAAGAGAGGCCCTGTGTACCATCCAGAGAGGCCACGGACAGC 345
      |||
Qy 453 ttccacacagaactccatccccgggtggcctcttgatcattaaagctgccacggcgaagg 512
      |||
Db 344 TTCCACACAGAACTCCATCCCCGGGTGGCCTCTGGATCATTAAGTCCACAGCGCAAGG 285
      |||
Qy 513 tcccacaggatgccttgaggggcgccactggctcagcaggaagcgacacgcctgcag 572
      |||
Db 284 TCCCACAGGATGCCCTGGAGGGCGCCACTGGCTCAGCGAAGAGCGACACCGCCTGCAG 225
      |||
Qy 573 gccatccgggatggaactccgcaagggggaccaccaaggaagcgtcctagaagaggggacccag 632
      |||
Db 224 GCCATCCGGGATGACTCCGCAAGGGGACCCACAAAGGACGCTCTAGAAGAGGGGACCGAG 165
      |||
Qy 633 agctctcccaactccaggctctccccgaagaccaccattactgtacatcctcagccc 692
      |||
Db 164 AGCTCTCTCCCACTCCAGGCTGTCCCCCGAAGACCCCACTTACTGTATCCTCAGGCC 105
      |||
Qy 693 tctcggcagctgtagggggtgggacggggagcacctgcctgttagcccccatcagacct 752
      |||
Db 104 TCTGGCAGCTGTAGGGGTGGGACCGGGGAGCACCTGCGCTGTAGCCCCCATCAGACCT 45
      |||
Qy 753 gccccaagcaccatatggaaataaagttcttc 785
      |||
Db 44 GCCCCAAGCACCATATGGAAATAAAGTCTCTTC 12
      |||

```

```

Insert Length: 1351      Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 461.
Location/Qualifiers
1. .593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1737911"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGGCGCGCCCAATTTTTTTTTTTT 3']".
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTT3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Ronaldo."
99 a 169 c 185 g 150 t

```

```
Query Match 49.5%; Score 388.2; DB 42; Length 593;
Best Local Similarity 99.2%; Pred. No. 8.1e-77;
Matches 390; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 393 aaggtaccagatggaggaagaagagccctgtgtaccatccagaaggccacggacagc 452
Db 404 AAGGTACCAGGAATGGAGGAAGAGAGGCCCTGTGTACCATCCAGAAGGCCACGACAGC 345
Qy 453 ttccacacagaactccatcccccgggtggtcttggatcatttaagctgccaagcgag 512
Db 344 TTCCACACAGAATCCATCCCGGGTGGCTTCTGGATCATTAAGCTGCCACGGCGGAGG 285
Qy 513 tccacacagatgacctggagggcgccactggctcagcgagaagcgacacgcctgcag 572
Db 284 TCCACACAGGATGCCCTGGAGGGCGGCCACTGTGCTCAGCGAGAGAGCAGACCGCTGCAG 225
Qy 573 gccatccggatgactccgcaagggaagccacacagaagcgtctctagaagaggacagc 632
Db 224 GCCATCCGGGATGACTCCGCAAGGGAGGCCACACAGGACGCTCTCTAGAGAGGGACCGAG 165
Qy 633 agctcctccactccagctgtccccccgaagacccacttactgtacatcctcagccc 692
Db 164 AGCTCTCCCACTCCAGCTGTCCCCCGGAAAGACCCACTTACTGTACATCCTCAGGCCC 105
Qy 693 tctcggcagctgtagggtggggacccggagcactgctgttagcccccacacacct 752
Db 104 TCTCGGAGCTGTAGGGGTGGGACCGGGAGACCTGCTGTAGCCCCCATCAGACCT 45
Qy 753 gcccaagcaccatggaaataaagtctttc 785
Db 44 GCCCAAGCACCATATGGAATAAAGTTCTTTC 12

RESULT 9
A1138943/c
LOCUS qd79a06.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1735666
DEFINITION 3', mRNA sequence.
ACCESSION A1138943
VERSION A1138943.1 GI:3644915
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 481)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2286288.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 979 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 456.
Location/Qualifiers
1. .481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:1735666"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
```

```
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGAGCGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
was constructed by one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 77 a 142 c 147 g 113 t 2 others
ORIGIN

Query Match 49.2%; Score 386.2; DB 42; Length 481;
Best Local Similarity 98.7%; Pred. No. 2.1e-76;
Matches 388; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 393 aaggtaccagatggaggaagaagagccctgtgtaccatccagaaggccacggacagc 452
Db 407 AAGGTACNAGATGGAGGAAGAGGCCCTGTGTACCATCCAGAAGGCCACGACAGC 348
Qy 453 ttccacacagaactccatcccccgggtggtcttggatcatttaagctgccaagcgag 512
Db 347 TTCCACACAGAATCATCNCCGGTGGCTTCTGGATCATTAAGCTGCCACGGCGGAGG 288
Qy 513 tccacacagatgacctggagggcgccactggctcagcgagaagcgacacgcctgcag 572
Db 287 TCCCAACAGGATGCCCTGGAGGGCGGCACCTGGCTCAGCGAGAAGCGACACCGCTGCAG 228
Qy 573 gccatccggatgactccgcaagggaagccacacagaagcgtcctagaagaggacagc 632
Db 227 GCCATCCGGATGACTCCGCAAGGGAGGCCACCAAGGACGCTCTAGAGAGGGACCGAG 168
Qy 633 agctcctccactccagctgtccccccgaagacccacttactgtacatcctcagggccc 692
Db 167 AGCTCTCCCACTCCAGGCTGTCCCCCGGAAAGACCCACTTACTGTACATCCTCAGGCCC 108
Qy 693 tctcggcagctgtagggtggggacccggagcactgctgttagcccccacacacct 752
Db 107 TCTCGGAGCTGTAGGGGTGGGACCGGGAGACCTGCTGTAGCCCCCATCAGACCT 48
Qy 753 gcccaagcaccatggaaataaagtctttc 785
Db 47 GCCCAAGCACCATATGGAATAAAGTTCTTTC 15

RESULT 10
A1652314/c
LOCUS A1652314 382 bp mRNA EST 17-DEC-1999
DEFINITION W60a03.x1 NCI-CGAP_Gc6 Homo sapiens cDNA clone IMAGE:2310028 3',
mRNA sequence.
ACCESSION A1652314
VERSION A1652314.1 GI:4736293
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 382)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 7, 1998 this sequence version replaced gi:3121269.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
```

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 462 Std Error: 0.00
Seq primer: -400P from Gibco.
Location/Qualifiers
1. 382

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2310028"
/clone_lib="NCI-CGAP_G66"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Plasmid DNA from the normalized library
NCI-CGAP_G64 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clones 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 59 a 103 c 131 g 89 t
ORIGIN

Query Match 46.7% Score 366.8; DB 49; Length 382;
Best Local Similarity 99.5%; Pred. No. 4e-72; Indels 0; Gaps 0;
Matches 368; Conservative 0; Mismatches 2;

```

QY 416 ggagggccctgtaccatccagaaggccagcagcttcacacagaactccatcccg 475
DB 382 GGAGGGCCCTGTGATCCTCCAGAGGCCAGGACAGCTTCCACACAGACTCCCTCCG 323
QY 476 ggtggccttgcattgaattgaattgcacagcggcggaggtccacagagatgccctggagg 535
DB 322 GGTGGCCTTCTGGATCATTAAGTGTGCACGCGGAGGTCCACAGGATGCCCTGGAGGG 263
QY 536 cggccactggtcagcagagcagcagcagcctgcagcagcctgcagcagcctgcagcagc 595
DB 262 CAGCAGCTGCTCAGCGAGAGCGACACCGCCCTGCAGGCCATCCGGATGGATCCGCGAA 203
QY 596 ggggaccccaagagcgtcttagaaggaggacagcagcagcctccactccagcagctgtc 655
DB 202 GGGGACCCACAGGACGCTCTAAAGAGGGGACCGAGAGCTCTCCACTCCAGGCTGTC 143
QY 656 ccccgaaagacccacttactgtacatctcagcagcctcctcagcagcctcctcagcagc 715
DB 142 CCCCCGAAAGACCCACTTACTGTACATCTCAGGCCCTCTCGGACGCTAGGGGTGGGG 83
QY 716 accggggagcacctgctgtagcccccatcagaccctgcacacccctgcacacacacata 775
DB 82 ACCGGGAGACACCTGCTGTAGCCCCCATCAGACCTGCCCCAAGCACCATTATGGAATA 23
QY 776 aagttcttc 785
DB 22 AAGTCTTTC 13

```

RESULT 11
AI377443/c
LOCUS
DEFINITION
IMAGE:2065471_3', mRNA sequence.
ACCESSION
AI377443
VERSION
AI377443.1 GI:4187296
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE

AUTHORS
TITLE

JOURNAL
COMMENT

Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 368)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jun 15, 1998 this sequence version replaced gl:3226479.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 772 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 312.

FEATURES

source

1. 368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2065471"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 56 a 97 c 123 g 92 t

Query Match 43.9% Score 345; DB 45; Length 368;
Best Local Similarity 100.0%; Pred. No. 2.8e-67;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 441 gccaggacagcttcacacagaactccatcccggtggtcctctggtatcattagctg 500
DB 368 GCACGGGACAGCTTCCACACAGACTCCATCCCGGCTTCTGGATCATTAAGTGTG 309
QY 501 ccacggcggaggtccacagcagatgccctggaggcggcactggcctcagcagagcga 560
DB 308 CCACGGCGGAGGTCCACAGGATGCCCTGGAGGGCGCCACTGGCTCAGCGAGAAGCGA 249
QY 561 caccgctgcaggccatccgggagtgactccgcaaggaggaccacacagagctcctagaa 620
DB 248 CACCGCTTCAGGCCATCCGGATGGACTCCGAGGGGACCCACAGGAGCTCTAGAA 189
QY 621 gaggggacagagctcctccactccaggctgtccccccgaagaccacttactgtac 680
DB 188 GAGGGGACCGAGAGTCTCTCCACTCCAGGCTGTCCCCCGGAAGACCCACTTACTGTAC 129
QY 681 atcctcagcctcctcggcagctgtagggtggggaccggggagcactgctgtagccc 740
DB 128 ATCTCAGGCCCTCTCGGACGCTGTAGGGGTGGGGAGCCGGGAGCACCTGCTGTAGCCC 69
QY 741 ccatacagacctgcccccaagcaccatattgaaataaagttcttctc 785
DB 68 CCATCAGACCTGCCCCAAGCACCATTATGGAATAAAGTCTTTC 24

```

RESULT 12

AA397836

LOCUS

DEFINITION

AA397836

440 bp mRNA

EST

16-MAY-1997

Soares testis_NHT Homo sapiens cDNA

clone IMAGE:728407

5' similar to UTR:G517093 G517093 HYPOPHYSICAL 39.2 KD PROTEIN. ;

AA397836

```

VERSION AA397836.1 GI:2050593
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 440)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
TITLE Unpublished (1997)
JOURNAL On Dec 30, 1996 this sequence version replaced gi:1528967.
COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 392.
FEATURES
source
1..440
/organism="Homo sapiens"
/db_xref="GDB:5925320"
/db_xref="taxon:9606"
/clone="IMAGE:728407"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAAGTGGGCGGCGCCCAATTTTTTTTTTT 3' ].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 98 a 142 c 120 g 80 t
ORIGIN
Query Match 42.7%; Score 335.4; DB 33; Length 440;
Best Local Similarity 91.9%; Pred. No. 4.1e-65;
Matches 376; Conservative 0; Mismatches 31; Indels 2; Gaps 2;

Qy 23 ccgggctgtgtctagcataaaggcgagccagcaagaaggcggtatgggagaagc 82
Db 1 CCGGGCTGTGCTCTAGCATAAAGCGG-CCGAGAAGATAGGGCGGGGTATGGGAGAAGC 59

Qy 83 ctccccactcccccaaggcgatctgtgtctgtctgtctgtctctctacct 142
Db 60 CTCGCCACCTGCCCGCGAA-CCGGCATCTGCTGCTGCTGCTCTCTACCCCT 118

Qy 143 ggtgatccctccgctgcagctctctatcatgatgctgagcccaagagagctcttggg 202
Db 119 GGTGATCCCTCCGCTCGAGCTCTATCCATGATGCTGACGCCCAAGAGAGCTCTTGGG 178

Qy 203 tctcacaggctccagagctactcccaaggcttcagcgactttctctgaagtgaaacct 262
Db 179 TCTCACAGGCTCCAGAGCTTCTCAAGGCTTCAGCGAGCTTTTCTCGAAAGGTAACT 238

Qy 263 gcttcggggatagacagcttattctctgcccccatggacttcggggcctccctgggaa 322
Db 239 GCTTCGGGGATAGACAGCTATTCTCTGCCCCCATGACTTCCGGGGGCTCCCTGGGAA 298

```

```

Qy 323 ctaccacaaaggagagaccagagcaccagctggggaaacaacaccctctccagccact 382
Db 299 CTACCACAAAGGAGAGACACAGGAGCACCAGCTGGGGAACACACCTCTCCAGCCACT 358

Qy 383 ccagatcgacaaagtaccaccaggatggaggagagagagggccctgtgtacc 431
Db 359 CCAGATCCACAAGATGACCGACACAACAGACAGAGAGGTGCTGATCTCC 407

RESULT 13
AA868111/c
LOCUS AA868111 466 bp mRNA EST 16-MAR-1998
DEFINITION ak31906.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1407610
3' similar to contains element MSRL repetitive element ;, mRNA
sequence.
ACCESSION AA868111
KEYWORDS AA868111.1 GI:2963556
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 466)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152384.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 465.
FEATURES
source
1..466
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1407610"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAAGTGGGCGGCGCCCAATTTTTTTTTTT 3' ].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 78 a 131 c 147 g 109 t 1 others
ORIGIN

Query Match 38.8%; Score 304.2; DB 39; Length 466;
Best Local Similarity 98.5%; Pred. No. 3.6e-58;
Matches 328; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 453 ttccacacagaactccatccccgggtggtctcttgatcatataagctgccacgcggagg 512
Db 342 TTTTACACAGAAGCTCCATCCCGGGTGGCTTCTGGATCATTAAGCTGCCACGGCGGAG 283

Qy 513 tccccaccaggtgctctggaggcgccactggctcagcgagagcgacacccgctcgag 572

```

```

|||||
Db 282 TCCACAGGATG-CCTGGAGGGCGGCACTGGCTCAGCGAGAAGCGACACCCCTGCAG 224
|||||
Qy 573 gccatccgggatgactcccgcaaggagaccacacaagagacgtcctagaagaggggaccggag 632
|||||
Db 223 GCCATCCGGGATGACTCCCGAAGGGGACCCACACAGGACGCTCTAGAAAGGGGACCGAG 164
|||||
Qy 633 agtctctccactccagctgtctcccccgaagaccacttactgtacatctctcagggccc 692
|||||
Db 163 AGCTCTCCCACTCCAGGCTGTCCCGCGAAGACCCACTTACTGTACATCTCTCAGG-CC 105
|||||
Qy 693 tctcggcagctgtagggtgagcagggagcacctgctgtagcccccatcagacct 752
|||||
Db 104 TCTCGGAGCTGTAGGGTGGGACCGGGAGCACCTGCTGTAGCCCCCATCAGACCT 45
|||||
Qy 753 qccccaagaccatattggaataaaagtcttttc 785
|||||
Db 44 GCCCCAAGCACCATATGGAATAAAGTTCTTTC 12

RESULT 14
AA854987/c
LOCUS
DEFINITION
aj53906.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394074
3', mRNA sequence.
ACCESSION
AA854987
VERSION
AA854987.1 GI:2942525
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 269)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 17, 1998 this sequence version replaced gi:1900909.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 956 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 214.
Location/Qualifiers
1. .269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394074"
/sex="male"
/lab_host="Soares_testis_NHT"
/Note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5]
TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
42 a 74 c 91 g 62 t

BASE COUNT
ORIGIN

```

```

Query Match 32.9% Score 258; DB 39; Length 269;
Best Local Similarity 100.0%; Pred. No. 5.7e-48;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 528 ctgagagggccactggctcagcgagaagcgacacgcctgcagggccatcccgatgga 587
|||||
Db 269 CTGAGGGCGGCCACTGGCTCAGCGAGAAGCGACACCGCCTGCAGGCCATCCGGGATGGA 210
|||||
Qy 588 ctccgcaaggggaccacacaagagcgtctctagaagaggggaccgagctctctccactcc 647
|||||
Db 209 CTCGCGAAGGGGACCCACAAAGGAGCGTCTTAGAAGAGGGGACCGAGAGCTCTCCCACTCC 150
|||||
Qy 648 aggtgtcccccgggaagaccacttactgtacatctctcagggccctctcggaagctgtag 707
|||||
Db 149 AGGCTGTCCCGCGAAGACCCACTTACTGTACATCTCTCAGGCCCTCTCGGAGCTGTAG 90
|||||
Qy 708 ggtgagggaccgggagcactgcctgtagccccatcagacctgccccaaagaccata 767
|||||
Db 89 GGGTGGGGACCGGGGAGCACCTGCTGTAGCCCCCATCAGACCTGCCCCAAGCACCATA 30
|||||
Qy 768 tggaataaagtcttttc 785
|||||
Db 29 TGGAAATAAAGTTCTTTC 12

RESULT 15
AI004529/c
LOCUS
DEFINITION
ot57q08.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1620926
3', mRNA sequence.
ACCESSION
AI004529
VERSION
AI004529.1 GI:3214039
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 257)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Feb 11, 1998 this sequence version replaced gi:2339722.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 965 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 220.
Location/Qualifiers
1. .257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1620926"
/sex="male"
/lab_host="Soares_testis_NHT"
/Note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5]
TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTTTTTT 3'].

```

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 41 a 67 c 88 g 61 t

ORIGIN

Query Match	31.3%;	Score 246;	DB 41;	Length 257;
Best Local Similarity	100.0%;	Pred. No. 2.6e-45;		
Matches 246;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	540	cactggtcagcgagaagcgacacgcgctgcaggccatccggatggactcgcgcaagggg	599	
DB	257	CACTGGCTCAGCGAGAGCGACACCGCCTGCGAGGCCATCCGGGATGGACTCGCAAGGGG	198	
QY	600	accacaaggagctctagaagagggaccgagagctcctccactccaggctgtccccc	659	
DB	197	ACCCACAAGGACGCTCTAGAGAGGGGACCGAGAGCTCTCCACTCCAGGCTGTCCCCC	138	
QY	660	cgaagaccacttactctacatcctcaggccctctcgacgtgtagggtgggaccg	719	
DB	137	CGAAGACCCCACTTACTGTATCTCATCTCAGGCCCTCTCGGCAGCTGTAGGGGTGGGACCG	78	
QY	720	gggagcacctgctgtagcccccatcagaccctgcccccaagcaccatatggaaataaagt	779	
DB	77	GGGAGCACCTGCTGTAGCCCCCATCAGACCCTGCCCCCAAGCACCATATGGAAATAAAGT	18	
QY	780	tctttc 785		
DB	17	TCTTTC 12		

Search completed: May 23, 2000, 22:03:55

Job time: 3993 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:03:55 ; Search time 3267.04 Seconds
(without alignments)
1058.267 Million cell updates/sec

Title: US-09-215-435-123

Perfect score: 853

Sequence: 1 ggaggatggcgagcagctct.....tttgcgaaaaaaaaaaaaa 853

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

Db 314 GTATACCCACAGAAGCAGATCAATTTGATGTGGTGCACAAAATGTGTGAGTTTCC 373
 QY 360 actaactgagcagctcatgagagaaatttggtagtcccggaacacacaaatagcgggattga 419
 Db 374 ACTAACTGAGCAGCTTCATGAGAAATTTGTTGATNCCGGAACACAAATAGCGGATTGA 433
 QY 420 tctccttagagacctatctttggcggtgcacagttcccttttaccatttggtagt 471
 Db 434 TCTCCTTAGAGACCTATCTTTGGCGTTG-CAGTTTCTTTTACCTTTGTGTGAGT 484

RESULT 3
 AA040832 514 bp mRNA EST 30-AUG-1996
 LOCUS ZK71704.1 Soares_pregnant_uterus_NbHPU Homo sapiens cdna clone
 DEFINITION IMAGE:485983.1, mRNA sequence.
 ACCESSION AA040832
 VERSION AA040832.1 GI: 517128
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 514)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilson,R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 396.
 FEATURES
 Location/Qualifiers
 1..514
 /organism="Homo sapiens"
 /db_xref="GDB:3759697"
 /db_xref="taxon:9606"
 /clone="IMAGE:485983"
 /clone_lib="Soares_pregnant_uterus_NbHPU"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /notes="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
 Site_2: Eco RI; 1st strand cdna was primed with a Not I -
 oligo(dT) primer [5'
 AACTGGAAGATTCGGCGCCGCTTTTTTTTTTTTTTTT 3']
 double-stranded cdna was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."

BASE COUNT 137 a 97 c 124 g 154 t 2 others
 ORIGIN

Query Match 50.7%; Score 432.2; DB 27; Length 514;
 Best Local Similarity 96.04; Pred. No. 1.9e-108;
 Matches 475; Conservative 0; Mismatches 15; Indels 5; Gaps 3;
 QY 4 ggaatggcgagcagctggaatgccagatgataaccgttttgcctacgatttgaatt 63
 Db 1 GGATGGCGGACGACTGTGAATGCCAGATGGATTAACCGTTTTTGTCTACGACGATTGTAAAT 60
 QY 64 gcttgtagtgcctcatttcaccatctccatgagcagctccattggcacagattc 123

Db 61 GCTTGTGCTAGCCATTTCCACCATCTACATGGCAGCCTCCATGGCAGACTTC 120
Qy 124 tggatgaatacgaagccaggtccagaaattccagtgatttgaaataaagacatctgg 183
Db 121 TGGTATGAATATCGAAGTCCAGTTCAGAAATTCAGTGTATTGATTAAGATCACTGG 180
Qy 184 gatgaattcattagatgaagcagatgaagaagactataatgatgcaccttttcgatac 243
Db 181 GATGAATTCATAGTATGAGCGCAGATGAAAGAGACTTATAATGATGACATTTTTCGATAC 240
Qy 244 aatgacagtgattgtgagacgtgtatcacaccatacccaaaaaca-tgcaattgta 302
Db 241 AATGGCAGTGGGATTGTGAGACGGTGTATCACCATTACCCAAAACATTGCAATTGGTA 300
Qy 303 tagccaccagaagagcagatgattgattgtgtcacaataatgtgtgatttcacact 362
Db 301 TAGCCACCAGAAAGGACAGATCATTTGATGTGTCACAAATGTGTGAGTTTCACACT 360
Qy 363 aactgagcagtcacagagaattgtgtgaccccggaacacacaatag--cgggattgat 420
Db 361 AACTGACAGTTCATGAGAAATTTGTTGATCCCGGAAACACAAATANGCGGATTGTAT 420
Qy 421 ctccatgagacctatttggcgtgacagttcccttttaccctttt--tgagtttagtt 478
Db 421 CTCCTTAGGACCTATCTTGGCGTGCAGTTCCTTTTACCTTTTGTGTGAGTTTAGGTT 480
Qy 479 tgatgtctttgggg 493
Db 481 TGATGTGNCRTTGGG 495
RESULT 4
AA310233
LOCUS EST181063 Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence. 19-APR-1997
DEFINITION AA310233
ACCESSION AA310233
VERSION AA310233.1 GI:1962560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 436)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.D., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudet,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
12140200
TITLE
JOURNAL Initial assessment of human gene diversity and expression patterns
MEDLINE based upon 83 million nucleotides of cDNA sequence
COMMENT Nature 377 (6547 Suppl), 3-174 (1995)
On Sep 12, 1996 this sequence version replaced gi:1394588.
Other_ESTs: THCL717196
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
I. 436
/organism="Homo sapiens"
/db_xref="ATCC (inhost):156284"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells V"
/cell_type="T-lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 130 a 83 c 105 g 116 t 2 others
ORIGIN
Query Match 48.9%; Score 417; DB 31; Length 436;
Best Local Similarity 98.6%; Pred. No. 2.7e-104;
Matches 420; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 ggaggatggcggcagctggaatgccagaatggataaccggtttgtctacagacatttga 60
Db 7 GGAGGATGGCGCAGCAGTCTGAATGCCAGANTGGATAACCGTTTGTCTACAGCATTTGTA 66
Qy 61 attgctgtgtccttagcctcatttccaccatctacatgcagcctccattggcacagac 120
Db 67 ATTGCTGTGTGCTTAGCCTCATTTCCACCATCTACATGGCAGCTCCATGGCAGAC 126
Qy 121 ttctggtatgaatatcgaagtcaggttccaggtaattccagtgatttgaataaaagcattc 180
Db 127 TTCTGGTATCAATATCGAAGTCCAGTTCAGAAATTTCCAGTATTGTAATAAAGCATC 186
Qy 181 tgggataaattcattagtgatgaagcagatgaagaagactataatgatgcaccttttcga 240
Db 187 TGGGATGAATTCATTCATAGTATGATGAGCGCAGATGAAAGACTTATAATGATGACACTTTTCGA 246
Qy 241 tacaatggcacagtgaggattgtgagacgtgtatcacaccatacccaaaaacatgcattgg 300
Db 247 TACAATGGCACAGTGGGATTGTGGAGACGGTGTATCACCATACCCAAAACATGCAATTGG 306
Qy 301 tatagccaccagaagaagacagagtcatttgatgtgtgcacaaaatgtgtgagtttcaca 360
Db 307 TATAGCCACCAGAAAGGACAGAGTCAATTTGATGTGTGTCACAAATGTGTGAGTTTCACA 366
Qy 361 ctaactgagcagtcactgagaaatttggatgcccggaacacacaatagcgggattgat 420
Db 367 CTAACTGAGCAGTTCATGGAGAAATTTGTTGATCCCGGAAACCAATAGCGGGATTGAT 426
Qy 421 ctctct 426
Db 427 CCTCCT 432
RESULT 5
AA451575/c
LOCUS AW451575
DEFINITION UI-H-B13-ald-f-11-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2736525 3', mRNA sequence.
ACCESSION AW451575
VERSION AW451575.1 GI:6992351
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 429)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On Mar 10, 1998 this sequence version replaced gi:2948561.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NC1-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bbrp/image/image.html The following repetitive elements were found in this cDNA sequence: 16-293, >MLTIF#LTR/MaLR Seq primer: M13 Forward
POLYA=Yes.

FEATURES	source
Location/Qualifiers	
i. 429	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:2736525"	
/clone_lib="NCI_CGAP_Sub5"	
/lab_host="DH10B (Life Technologies)"	
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; NCI_CGAP_Sub5 is a subtracted library derived from NCI_CGAP_Sub4. The NCI_CGAP_Sub5 library had 3 million recombinants. A single-stranded DNA preparation of NCI_CGAP_Sub4 was used as a tracer in a subtractive hybridization with a driver comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3693, 3798-3803 (IMAGE Clonoids 1323376-1323911, 1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids 1323912-1325831, 1471368-1472803, 1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351). (10% of the driver population), plus a pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of the driver population), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the driver population). Subtraction was performed as previously described [Bonaldi, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806. TAG LIB=NCI_CGAP_Pr22 TAG TISSUE=prostate TAG_SEQ=AAGTG"	

BASE COUNT	123 a	87 c	104 g	115 t	ORIGIN
	iAG_SEQ=AAAGG				
Query Match	48.2%; Score 411.4; DB 81; Length 429;				
Best Local Similarity	98.6%; Pred. No. 9.4e-103;				
Matches 415; Conservative	0; Mismatches 6; Indels 0; Gaps 0;				
Qy 433	tatcttggcgttcgcagttccctttaccctttcttgatgttagttgagttgctttggg	492			
Db 429	TATCTTTGGCGTTCGCAGTTCCTTTTACCCTTTCTGAGTTTAGTTCGATGGTCTTGGG	370			
Qy 493	gctttgatcggacttttctgtctgcatctgcgaagcttatatccacattgcacagggc	552			
Db 369	GCTTTGATCGGACPTTGTGCTTGCATTTGCCGAAGCTTATATCCACCATTCGCACGGGC	310			
Qy 553	attctcatctccttgcagataccatctgctggaagtcacaggccacatggagagtgctcctgt	612			

Db	309	ATTCTCCATCTCCTTGCGAGATACCATGCTGTGAAGTCCAGGCCACATGAGAGGTGTCTGT	250
Qy	613	gtagatgctccagctgaaatcccccaagctaaagctccccaaactcacagccaaacatctttcca	672
Db	249	GTAGATGCTCCAGCTGAAATCCCAAGCTTAAGCTCCCAACTGACAGCCAAACATATTCCA	190
Qy	673	gcattgtgtggagccatccttgatgtccagccttaacaaagccttcagagagacttcagcc	732
Db	189	GCCATGTGTGGAGCCATCCTTGATGTGTCAGCCCTTAACAGCCCTTCAGAGACTTCAGCC	130
Qy	733	acagctattcttactacatcctgttgagactctaataaagaacaaactagctgagccc	792
Db	129	ACAGTATTATCTTACTACATCCTTGTGAGACTCTAATAAAGAACCACTAGCTGAGCCC	70
Qy	793	aatcaacctatggaaactgatgaataaataaagtgtgtttgctgagaaaaa	852
Db	69	AATCAACCTATGGAACCTAGATAAATAAATGAATGTGTGTCGCCGCTAAAAAAA	10
Qy	853	a 853	
Db	9	a 9	
RESULT	6		
W01548			
LOCUS	W01548	501 bp	EST
DEFINITION	za34c10.r1 Soares fetal liver spleen 1NFLS Homo sapiens CDNA clone	IMAGE:294450	5', mRNA sequence.
ACCESSION	W01548		
VERSION	W01548.1	GI:1273528	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 501)		
	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,		
	Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,		
	Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,		
	Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and		
	Wilson, R.		
TITLE	The WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	On Apr 14, 1993 this sequence version replaced gi:837457.		
	Contact: Wilton RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LLNL ; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Seq primer: mob.REGA+ET		
	High quality sequence stop: 370.		
FEATURES	Location/Qualifiers		
source	1..501		
	/organism="Homo sapiens"		
	/db_xref="GDB:1239371"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:294450"		
	/clone_lib="Soares fetal liver spleen 1NFLS"		
	/sex="male"		
	/dev_stage="20 week-post conception fetus"		
	/lab_host="DH10B (ampicillin resistant)"		
	/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)"		
	with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;		
	1st strand cDNA was primed with a Pac I - oligo(dT) primer		
	[5', AACTGGAAGATTAATTAAGATCTTTTTTTTTTTTTTTT 3'],		
	double-stranded cDNA was ligated to Eco RI adaptors		
	(Pharmacia), digested with Pac I and cloned into the Pac I		
	and Eco RI sites of the modified pT7T3 vector. Library		
	went through one round of normalization. Library		
	constructed by Bento Soares and M.Fatima Bonaldo."		

BASE COUNT	135 a	94 c	120 g	151 t	1 others
Query Match	47.7%	Score 406.8	DB 25	Length 501	
Best Local Similarity	95.8%	Pred. No. 1.8e-101			
Matches	482	Conservative 0	Mismatches 13	Indels 8	Gaps 6
QY	11	cgagcagctgaatgcagaaatggataacccgttttgctacagcattgttaattgctgtg 70			
Db	1	CGAGCAGCTGTAATGCCAGATGGATAACCGTTTGCTACAG-AATTGTAATGCTGTG 59			
QY	71	tgtctagacctcaattccaccattccatcagcagctccattggcagacactctggtatg 130			
Db	60	TGCTTAGCCCTCAATTTCCACCATTCTACATGGCAGCCTCCATTGGCACACACTTCGTGGTATG 119			
QY	131	aatatcgaagtcaggttcaagaataattccagtgatttgaataaaagcatctgggatgaat 190			
Db	120	AATATCGAAGTCCAGTTCACGAATAATCCAGTGATTGTAATAAAGCATCTGGGATGAAT 179			
QY	191	tcattagatgaagcagatgaagaagcattataatgatgcaccttttcgcatacaatggca 250			
Db	180	TCATTAGTGTATGAGGCAGATGAAAGACTTATAATGATGCACTTTTTCGATACAAATGGCA 239			
QY	251	cagtggggattgtgagacggtgtatcaccatacccaaaaacatgcatggtatagccac 310			
Db	240	CAGTGGGATTGTGGAGAGGGTATACACCATACCCAAAACATGCAATGGTATAGCCCCAC 299			
QY	311	cagaaagcagagatcatttgatggtgcacaaatgtgtgagtttcacactaactgagc 370			
Db	300	CACGAAG-GACAGTCATTGTGATGGTGCACAAATGTGTGAGTTTCACACTAACTGAGC 358			
QY	371	agttcatggaataattgtgtgattccgggaacacacataagcgga-ttgatctccttag 429			
Db	359	AGTTCATGGGAAATTTGTATCCCGGAACACACATAACGGGGATTTGATCTCTTAGG 418			
QY	430	acctatotttgccgttgcca--gttccttttacctttgtgagttta-ggtttgatgtgc 486			
Db	419	ACCTATCTTTGGGCGTTGTCNAGGTTCTTTTACCTTTTGTGAGTTTAGGGTTGATGTCG 478			
QY	487	ttt--ggggcctttgatcgagactt 507			
Db	479	TTTGGGGGGCTTTGATCGGACTT 501			
RESULT 7					
AA150665					
LOCUS	AA150665	439 bp	mRNA	EST	10-DEC-1996
DEFINITION	z143f05.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:504705 5', mRNA sequence.				
ACCESSION	AA150665				
VERSION	AA150665.1	GI:1722177			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 439) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Roeding,T., Soares,M., Tan,F., Trevisan,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.				
TITLE	The WashU-Merck EST Project				
JOURNAL	Unpublished (1995)				
COMMENT	On May 5, 1995 this sequence version replaced gi:798266. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the				

Qy	419	atctcct 425
Db	441	ATCTCCT 447
RESULT 9		
AI183646/c		
LOCUS		
DEFINITION 388 bp mRNA EST 10-NOV-1998		
qf30F09.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1751561		
3', mRNA sequence.		
ACCESSION AI183646		
VERSION AI183646.1 GI:3734284		
KEYWORDS		
SOURCE		
ORGANISM		
Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
Eutheria; Primates; Catarrhini; Hominidae; Homo.		
1 (bases 1 to 388)		
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
NATIONAL CANCER INSTITUTE, Cancer Genome Anatomy Project (CGAP),		
Tumor Gene Index		
JOURNAL		
COMMENT		
On Jan 19, 1998 this sequence version replaced gi:2151890.		
Contact: Robert Strausberg, Ph.D.		
Tel: (301) 496-1550		
Email: Robert_Strausberg@nih.gov		
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima		
Bonaldo, Ph.D.		
CDNA Library Arrayed by: Greg Lennon, Ph.D.		
DNA Sequencing by: Washington University Genome Sequencing Center		
Clone distribution: NCI-CGAP clone distribution information can be		
found through the I.M.A.G.E. Consortium/LLNL at:		
www-bio.llnl.gov/bbrp/image/image.html		
Insert Length: 531 Std Error: 0.00		
Seq primer: -400P from Gibco		
High quality sequence stop: 376.		
Location/Qualifiers		
1..388		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="IMAGE:1751561"		
/clone_lib="Soares_testis_NHT"		
/sex="male"		
/lab_host="DH10B"		
/note="vector: p7T3D-Pac (Pharmacia) with a modified		
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA		
was prepared from mRNA obtained from Clontech		
Laboratories, Inc., and primed with a Not I - oligo(dT)		
primer [5'		
TGTTACCAGTCGAAGTGGAGCGGCCGCACATTTTTTTTTTTTTT 3'].		
Double-stranded cDNA was ligated to Eco RI adaptors		
(Pharmacia), digested with Not I and cloned into the Not I		
and Eco RI sites of the modified p7T3 vector. Library		
went through one round of normalization to Cot5, and was		
constructed by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT 110 a 82 c 98 g 98 t		
ORIGIN		
Query Match 44.4%; Score 378.4; DB 43; Length 388;		
Best Local Similarity 99.7%; Pred. No. 1.1e-93;		
Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy 460 ccttttgaggttaggttgcattgcttggcgcttgatcggaacttgtctgcaatt 519		
Db 388 CCTTTGTGAGTTAGGTTGCTATGCTTTGGGGCTTTGATCGGACTTTGCTTGCATT 329		
Qy 520 tgccagaagcttatccccaccattgccacggcgattcttcattctcttcgcagataccatg 579		
Db 328 TGCCGAAAGCTTATATCCCACCATTGCCACGGGCAATCTCCATCTCCTTGCAATACCATG 269		
Qy 580 ctgtgaagtccaggccacatgagggtgctctgtgatagctccagctgaaatcccgaagc 639		

```

|||||
268 CTGTGAAGTCCAGGCACATGAGGTTGCTCTGTGTAGATGCTCCAGCTGAATCCCAAGC 209
|||||
640 taagctcccaactgacagcaacacatcttccagccatgtgtggggccatcctgagatg 699
|||||
208 TAAGCTCCCAACTGACAGCAACATCATTTCCAGCCATGTGTGGGACCATCTCGATGCT 149
|||||
700 ccagccttaaacagccttcagaggaacttcagcaacagcattatcttactacatcctgt 759
|||||
148 CCAGCCTTAACAAGCCTTACAGAGACTTCAGCCACAGCTATTATCTACTACATCCTTGT 89
|||||
760 gagactctaataagaacaaactagctgagcccaatcaaccctatggaactgataagaata 819
|||||
88 GAGACICTATAAAGAACCAACTAGCTGAGCCCAATCAACCTATGCACTGATAGAAATA 29
|||||
820 aatgaattgtgttttgoy 839
|||||
28 AAATGAATTGTTGTTTGTG 9
|||||

RESULT 10
AW449183/c
LOCUS
DEFINITION
  UI-H-BI3-akg-e-03-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone
  IMAGE:2734180 3', mRNA sequence.
ACCESSION
  AW449183
VERSION
  AW449183.1 GI:6989959
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 395)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  On Jan 6, 2000 this sequence version replaced gi:6675921.
  Contact: Robert Strausberg, Ph.D.
  Tel: (301) 496-1550
  Email: Robert.Strausberg@nih.gov
  The sequence contained an oligo-dT track that was present in the
  oligonucleotide that was used to prime the synthesis of first
  strand cDNA and therefore this may represent a bonafide poly A
  tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
  NCI-CGAP clone distribution information can be found through the
  I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html The following repetitive
  elements were found in this cDNA sequence: 16-293, >MLT1F#LTR/MaLR
  Seq primer: M13 Forward
  POLYA=Yes.
  Location/Qualifiers
    1..395
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2734180"
      /clone_lib="NCI_CGAP_Sub5"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT73D-Pac (Pharmacia) with a modified
      polylinker; Site_1: Not 1; Site_2: Eco RI; NCI_CGAP_Sub5
      is a subtracted library derived from NCI_CGAP_Sub4. The
      NCI_CGAP_Sub5 library had 3 million recombinants. A
      single-stranded DNA preparation of NCI_CGAP_Sub4 was used
      as a tracer in a subtractive hybridization with a driver
      comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
      3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
      1322376-1323911, 1456008-1456775, 1500552-1502855);
      NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
      (IMAGE Clones) 1323912-1325831, 1471368-1472903,
      1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
      3851-3854 (IMAGE Clones)
      1414920-1417991, 1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM
      3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)

```

```

1257096-1258631,1469064-1470983, 1475592-1476743);
  NCI_CGAP_Pt22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
  (IMAGE Clones) 985608-986759,1101192-1101959,
  1217928-1220615); NCI_CGAP_Col0 pool 1 LLAM 2644-2653,
  2871-2872 (IMAGE Clones)
  1057416-1061255,1144584-1145351). (10% of the driver
  population), plus a pool of 3,840 arrayed clones from
  NCI_CGAP_Sub1 (IMAGE Clones) 2708616-2710535) and
  NCI_CGAP_Sub2 (IMAGE Clones) 2710536-2712455) (10% of
  the driver population), plus a pool of 11,136 clones from
  NCI_CGAP_Sub3 (IMAGE Clones) 2712456-2723531) (10% of the
  driver population), plus a pool of 5,472 clones from
  NCI_CGAP_Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
  driver population). Subtraction was performed as
  previously described [Bonaldo, Lennon & Soares (1996)]:
  Normalization and Subtraction: Two Approaches To
  Facilitate Gene Discovery. Genome Research 6, 791-806.
  TAG_LIB=NCI_CGAP_Lu19
  TAG_TISSUE=lung
  TAG_SEQ=GACAGC"

BASE COUNT      107 a      81 c      96 g      111 t
ORIGIN

Query Match      43.9%; Score 374.2; DB 81; Length 395;
Best Local Similarity 97.9%; Pred. No. 1.6e-92;
Matches 379; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 467 tgagtttagtttgatgtgcttttggggctttgatcggaacttgccttgcatttcgcaa 526
  |||||||
Db 395 TGAGTTTAGTTTGTATGTCTTTGGGGCTTTGATCGGACTTTTGTCTTCATTTGCCGAA 336
  |||||||
QY 527 gcttatatccaccattgccacggccattctccatctcttcagataccatgctgtgaa 586
  |||||||
Db 335 GCTATATCCCACTATGCGACGGGCATTCCTCCATCTCCTTCGAGATACATGCTGTGAA 276
  |||||||
QY 587 gtccaggccacatggaggtgctgtgtgtagtgcctcagctgaaatcccaagctaagctc 646
  |||||||
Db 275 GTCCAGGCCACATGGAGGTGCTCTGTGTAGATGCTCCAGCTGAAATCCCAAGCTAAGCTC 216
  |||||||
QY 647 ccaactgacagccaaacatcatttccagccatgtgtgggagccatcctgcagctc 706
  |||||||
Db 215 CCAACTTGACAGCCCAACATCATTTCCAGCCATGTGTGGGAGCCATCTCGATGTCCAGCCT 156
  |||||||
QY 707 taacaagccttcagaggaacttcagccacagcattattcttactacatccttgcagactc 766
  |||||||
Db 155 TAACAAGCCTTCAGATGACTTCAGCCACAGCTATTATCTTACTACATCCTTGTGAGACTC 96
  |||||||
QY 767 taataagaaccaactagctagtgagcccaatcaacctatggaactgataagaataaata 826
  |||||||
Db 95 TAATAAAGAACCAACTAGCTAGGAGCCCAATCAACCTATGGAAGCTGGTAGAATAAATAATGAA 36
  |||||||
QY 827 ttgttttttgcgaaaaaataaaaaaa 853
  |||||||
Db 35 TTGTTGTTTGTGCGCGCTAAATAAAAAA 9
  |||||||

RESULT 11
AA769827/c
LOCUS
DEFINITION
  ah81c02.s1 Soares_testis_NHT Homo sapiens cDNA clone 1322018 3',
  mRNA sequence.
ACCESSION
  AA769827
VERSION
  AA769827.1 GI:2821065
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 404)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index

```


Unpublished (1997)
On Jan 9, 1998 this sequence version replaced gi:936388.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```

Insert Length: 914      Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 395.

FEATURES
    source
        1..404
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="1322018"
            /clone_lib="Soares_testis_NHT"
            /sex="male"
            /lab_host="DH10B"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
            was prepared from mRNA obtained from Clontech
            Laboratories, Inc., and primed with a Not I - oligo(dT)
            primer [5].
            TGTTACCAATCTGAAGTGGGCGCGCCAAATTTTTTTTTT 3' ].
            Double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the Not I
            and Eco RI sites of the modified pT7T3 vector. Library
            went through one round of normalization to Cot5, and was
            constructed by Bento Soares and M. Fatima Bonaldo."
        106 a      79 c      96 q      123 t

BASE COUNT

```

Query Match	43.8%;	Score	373.8;	DB	38;	Length	404;
Best Local Similarity	98.2%;	Pred. No.	2e-92;				
Matches	378;	Conservative	0;	Mismatches	7;	Indels	0;
Gaps	0;						
QY	469	agtttaggttgatgtgctttggggccttggatcgaccttgtcttgcattgccgaagc	528				
DB	404	AGTTTAGGTTTGATGTGCTTTGGGCGTTTGATCGACTTTGTGCTTGCATTTGCCGAAGC	345				
QY	529	ttatatccaccattgccaggggcatcttccatctccttgcagataccatgctgtgaagt	588				
DB	344	TTATATCCCACCAATTTCCAGGGCATTTCTCCATCTCTTGCGATACCATCTGTGAAGT	285				
QY	589	ccaggccacatggagggtgctctgtgtagatgctccagctgaatcccaagctagctccc	648				
DB	284	CCAGGCCACATGGAGGTGCTCTGTGTAGTGTCCAGTGAATCCCAAGCTAAGCTCCC	225				
QY	649	aactgcagcgcacacatcaittccagcccatgtgtggagccatcctgcatgtccagcccta	708				
DB	224	AACGTACAGCCCAACATCATTTCCAGCATGTGTGGGAGCCATCTCTGGATGTCCAGCCTTA	165				
QY	709	acaagccttcagaggacttcagccacagctattatcttactacatccttgtgagactcta	768				
DB	164	ACAAGCCTTCAGAGGACTTCAGCCACAGCTATTATCTTACTACATCCTTGTGTGACACTA	105				
QY	769	ataagaaaccaactagctgagcccaatcaacctatggaactgatagaataaaaatgaatt	828				
DB	104	ATRAAGAACCACCTAGCTGAGCCCAATCAACTATGGAACCTATGAGAATAAAAATGAATT	45				
QY	829	gttgtttttgcgaaaaaataaaaaaa	853				
DB	44	GTGTGTTTTGTGCCCTTAAAAAATAA	20				
RESULT	12						

AA027882 456 bp mRNA EST 09-MAY-1997
LOCUS zk0507.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:469668 5', mRNA sequence.
ACCESSION AA027882
VERSION GI:1493979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 456)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., Dubuca,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kubaca,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.,
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On Apr 14, 1993 this sequence version replaced gi:716759.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 567 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 312.
Location/Qualifiers
1. 456
/organism="Homo sapiens"
/db_xref="GDB:3755670"
/db_xref="taxon:9606"
/clone="IMAGE:469668"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3-Pac; Site:1: Not I ;
Site:2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AATCGAAGAAATTCGGCGCGCCGCTTTTTTTTTTTTTTTT 3']/
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
BASE COUNT 130 a 91 c 109 g 124 t 2 others
ORIGIN
Query Match 43.4%; Score 370.6; DB 27; Length 456;
Best Local Similarity 98.3%; Pred. No. 1.6e-91;
Matches 395; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
Qy 4 ggaatggcgagcactgcgaatgccagaaatgataacgcgttttgcacagattgtaatt 63
|||||
Db 1 GGATGGCGGAGCAGCTGGAATGCCAGAAATGATTAACCGTTTTTGTCTACAGCAATTGTAATT 60
|||||
Qy 64 gcttggtgcttagcctcatttcaccatctcacatggcagcctcattggcacagacttc 123
|||||
Db 61 GCTTGTGTGCTTAGCCTCATTTCCACCATCTACATGGCAGCGCTTCATTGGCACAGACTTC 120
|||||
Qy 124 tggatgaatatcgaagtccagttccaagaaataatccacgtgattggaataaagcatctgg 183
|||||
Db 121 TGGTATGAATATCAAGTCCAGTTCAGAAATTCACGTGATTTGGAATAAAGCATCTGG 180
|||||

RESULT 12

RESULT 15					
AA103099	AA103099	473 bp	mRNA	EST	29-OCT-1996
LOCUS					
DEFINITION	mol7e03.r1	Life Tech mouse embryo 13 5dpc	10666014	Mus muscul	
		cDNA clone IMAGE:553852 5'		mRNA sequence.	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:04:01 ; Search time 3267.04 Seconds
(without alignments)
1024.769 Million cell updates/sec

Title: US-09-215-435-124
Perfect score: 826
Sequence: 1 aaccagagtgcccatgggt.....atcatcaaaaaaaaaa 826

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues 9714632
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result % Query
SUMMARIES

RESULT 2
 AI816715/c 547 bp mRNA EST 21-DEC-1999
 LOCUS wj43008.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2403559 3'
 DEFINITION similar to SW:D2.ONCVO P54187 D2 PROTEIN ; mRNA sequence.
 ACCESSION AI816715
 VERSION AI816715.1 GI:5435794
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 547)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 18, 1998 this sequence version replaced gi:3138649.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 903 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 453.

FEATURES

source
 1. .547
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2405559"
 /clone_lib="NCI-CGAP_Lu19"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT7N3 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."
 102 a 131 c 160 g 154 t

Query Match 61.8%; Score 510.2; DB 61; Length 547;
 Best Local Similarity 98.3%; Pred. No. 2.1e-131;
 Matches 526; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

Qy 273 ggcgcaacctatctcgtggtgagtcagatgccctagcagacagacccaga 332
 Db 547 GCGCAACCTATATCTCTGTGTGTGATCCAGATGCCCTAGCAGACGACCCAGA 488
 Qy 333 cagagattctggagacattgctggtgaacagatatcaaggcgccgacctgaagaaagg 392
 Db 487 CAGAGATTCTGAGACATTGGCTGTTACAGATATACAGGGCGCGGACCTGAGGAAGG 428
 Qy 393 aagattcaggccaggagttatcagctaccaggtccctccccaccgccacagtggtgc 452
 Db 427 AAGATTGAGGCGCAGGAGTTATACGCTTACCAGGCTCCCTCCCGCACACAGTGGC 368

Qy 453 ttccatcgctaccagttctctgtctctatcttcaggaggaaggtcatctctctctcc 512
 Db 367 TTCCATCGCTACCACTTCTTGTCTATCTTCAGGAAGAAAGTCTCTCTCTCTCC 308
 Qy 513 aaggaatacaaaactcgcaggtccttggaaaaatggacagattcttgaacgtttccacctg 572
 Db 307 AAGGAAAAACAAACTCGAGGCTCTTGGAAAAATGGACAGATTTTTCAGCGCTTCCACTTG 248
 Qy 573 ggcgaacctgaagaagcaccaggttcattgacccaggaactaccaggaactcacaacctc 632
 Db 247 GCGAACCCTGAAGCAAGCACCAGTTCATGATGCCAGAACTACCCAGACTCACAACCTTC 188
 Qy 633 caggtctccagagaagggccagcgagccagcaaaaacccagcgagagatagctgc 692
 Db 187 CAGGCTCCAGAGGAAGGGCCAGCGAGCCCRAGCAC-NAACACAGCGGAGAGTAGCTGCC 129
 Qy 693 tgcctagatagcgggtttgcccacatccggcgatgtgcccacactgccacacacgagatgt 752
 Db 128 TGCTAGATAGCGGCTTTGCCATCGGGCATGTGGCCACACTGCTCAGCACCGAGCATGT 69
 Qy 753 gggtatgaaacccctctggatcacagaacccctctcttccaaataaaaaaaa 807
 Db 68 GGGTATGGAACCCCTCTGGATACAGAACCCCTCTCTTCCAAATTAATAAAAAA 14

RESULT 3

AI857902/c
 LOCUS wj68a01.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407944 3'
 DEFINITION similar to SW:D2.ONCVO P54187 D2 PROTEIN ; mRNA sequence.
 ACCESSION AI857902
 VERSION AI857902.1 GI:5511518
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 547)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Feb 18, 1999 this sequence version replaced gi:4314810.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 752 Std Error: 0.00

Seq primer: -400P from Gibco
 High quality sequence stop: 450.

FEATURES

source
 1. .547
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2407944"
 /clone_lib="NCI-CGAP_Lu19"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to

AA688029
 VERSION AA688029.1 GI:2674935
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 511)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap/>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802277.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 879 Std Error: 0.00
 Seq primer: -40M13 fwd. ET from Amersham
 High quality sequence stop: 419.

FEATURES

source
 Location/Qualifiers
 1..511
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1220398"
 /clone_lib="NCI-CGAP-Pr22"
 /sex="male"
 /tissue_type="normal prostate"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
 with a modified polylinker; 1st strand cDNA was prepared
 from normal prostate bulk tissue, and was then primed with
 a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT7T3 vector. Library is normalized, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 92 a 118 c 155 g 146 t

BASE COUNT 92 a 118 c 155 g 146 t
 ORIGIN

Query Match 58.7%; Score 484.8; DB 37; Length 511;
 Best Local Similarity 98.4%; Pred. No. 2.3e-124;
 Matches 500; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Qy 300 gatccagatgcccctagcagacgagaccacacagagattcttgagacattggtgta 359
 Db 511 GATCCAGATGCCCTAGCAGAGCAGAGACCCACAGAGATTCTGGAGACATTGGCTGTA 452
 Qy 360 acagatatcaaggcgccgacctgaagaagggaagattccgagggccagaggttatcgcc 419
 Db 451 ACAGATATCAAGGGCGCGACCTGAAGGAAGGAAGATTTCAGGGCCAGGAGTTATCAGCC 392
 Qy 420 taccaggctccctccacccgacagtggtcttcctccatcgctaccagttcttcttat 479
 Db 391 TACCAGGCTCCCTCCCAACCGGACACAGTGGCTTCCATCGCTACCAAGTCTTTGTCAT 332
 Qy 480 cttcaggaggaagggttcattctctctcttcctccaggaacacacacacacacacacac 539
 Db 331 CTTCAGGAGGAAGATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 272
 Qy 540 aaaaatggacagattcttgaacccgtttccacctggcggaacctgaagcagcaccagttg 599
 Db 271 AAAATGGACAGATTCTTGAACCGCTTCCACCTGGCGGAACCTGAAGCAAGCACCAGTTC 212

Qy 600 atgaccagaactaccagactaccacacccctccaggtccagagaaagggccagcgag 659
 Db 211 ATGACCCAGAACTACCAGGACTACCACCTCCAGGCTCCAGAGGAAGGCCACGACG 152
 Qy 660 cccaagcacaaaaaccagggcgagatagctcctgctagatagccggcttgccatccgg 719
 Db 151 CCCAAGCAC-AAAACCAGGGCGAGATAGTCTGCTAGATAGCCGCTTTGCCATCCGG 93
 Qy 720 gcatgtggccacactgccaccaccgacgatgtgggtatggaaacccctctggatacaga 779
 Db 92 GCATGTGGCCACACTGCTCACCACCGACGATGGGTATGAACCCCTCTGGATACAGA 33
 Qy 780 accctctttttccaaataaaaaaaa 807
 Db 32 ACCCTCTCTTTTCCAAATTAATAAAAAA 5

RESULT 6

AA192427/c
 LOCUS

DEFINITION ZQ01g08.s1 Stratagene muscle 937209 Homo sapiens cDNA clone
 IMAGE:628478 3' similar to SW:PBP-RAT F31044
 PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN ; mRNA sequence.

ACCESSION AA192427

VERSION AA192427.1 GI:1781834

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 457)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)

JOURNAL

COMMENT

On May 8, 1995 this sequence version replaced gi:801531.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 658 Std Error: 0.00

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 379.

Location/Qualifiers

1..457

/organism="Homo sapiens"

/db_xref="GDB:5048712"

/db_xref="taxon:9606"

/clone="IMAGE:628478"

/clone_lib="Stratagene muscle 937209"

/tissue_type="muscle"

/dev_stage="adult"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: skeletal muscle; Vector: pBluescript SK-;

Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally.

Primer: Oligo dT. Skeletal muscle from patient with

malignant hyperthermia. Average insert size: 1.0 kb;

Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG

3' -3' adaptor sequence: 5' CTCAGATTTTTTTTTTTTTTTT 3'

BASE COUNT 84 a 101 c 140 g 131 t 1 others

ORIGIN

Query Match 51.6%; Score 426; DB 29; Length 457;

Best Local Similarity 97.4%; Pred. No. 5e-108;

Matches 443; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 569 Std Error: 0.00
Seq Primer: mob.REGA+ET
High quality sequence stop: 434.

Location/Qualifiers

1. 452

source

/organism="Homo sapiens"
/db_xref="GDB:1259475"
/db_xref="taxon:9606"
/clone IMAGE:321817"
/clone_lib="Soares_parathyroid tumor NBHPA"

TITLE Identification of 4370 expressed sequence tags from a 3'-end-specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization

JOURNAL Genome Res. 6 (1), 35-42 (1996)

MEDLINE 96276048

COMMENT On Mar 10, 1998 this sequence version replaced gi:2949496.
Contact: Valle G.
CRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
<http://grup.bio.unipd.it>.

FEATURES
source
1. 423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s4000032A04"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/note="Vector: pCDNII (Invitrogen); Site_1: BstXI; Site_2: NotI; The library was constructed by G. Lanfranchi. This library is not subtracted nor normalized. The first strand cDNA was primed with a biotinylated oligo-dT-NotI primer
(5'-biotin-AACCCGGCTCGAGCGCGCTTTT-3'). The ds cDNA was sonicated and size-selected in the range 350-550 bp. The 3' specific fragments were selected by streptavidin coated magnetic beads, ligated to non-palindromic BstXI adapters, NotI digested and directionally cloned into BstXI-NotI cut pCDNII vector."

BASE COUNT 114 a 130 c 103 g 76 t

ORIGIN

Query Match 49.1%; Score 405.2; DB 50; Length 423;
Best Local Similarity 98.1%; Pred. No. 3e-102;
Matches 410; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 349 attgctgtgatacatatcaaggcgccgacacagtgagtgcttccatgcctaccagt 408
Db 1 ATTGGCTGTAAACAGATATCAAGGGCGCGCAGCTGAAGGAAGGAAAGATTTCAGGGCCAGG 60

Qy 409 agttatcagctaccagctccctccacacgacacagtgagtgcttccatgcctaccagt 468
Db 61 agttatcagctaccagctccctccacacgacacagtgagtgcttccatgcctaccagt 120

Qy 469 tcttctgtctatcttcaggaaggaaggtcatctctctcttccacaggaacaaacac 528
Db 121 TCTTTGTCTATCTTCAGGAAGGAAGTCACTCTCTCTTCCCAAGGAACAAACATC 180

Qy 529 gaggctcttgaaatgacagattctgaaccgtttccacctggcgacactgaagcaa 588
Db 181 CAGGCTCTTGGAAATGACACAGATTCTCAACCGTTTCACCTGGCGCAACCTGAAGCAA 240

Qy 589 gcaccagttcatcaccagactaccaggactaccacacacacacacacacacacac 648
Db 241 GCACCCAGTTATGACCCAGAACTACCAAGGACTACCAACCCCTCCAGGCTCCAGAGAAA 300

Qy 649 gggccagcgagcccaagcacaaaaccaggcgagatagctgcctgctagatagccgct 708
Db 301 GGGCCAGCGAGCCCAAGCACAAAACCCAGCGGAGATAGTGCCTGCTAGTAGCGCGCT 360

Qy 709 ttgccatcgggcatgtggccacactgccaccacccagcagatgtgggtatgaaacccc 766
Db 361 TTGCCATCGGGCATGTAGGCACACTGCGCCACCCAGCAGATGTGGGTATGACCCCC 418

RESULT 9
AA854779/c 412 bp mRNA EST 04-JAN-1999
LOCUS
DEFINITION aj77b05.sl Soares parathyroid tumor_NbHPA Homo sapiens cDNA clone
IMAGE:1402449 3' similar to SW:PBP_RAT P31044

ACCESSION AA854779

VERSION AA854779.1 GI:2942317

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 412)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
On Jan 17, 1998 this sequence version replaced gi:2043553.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043553.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

FEATURES
source
1. 412
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1402449"
/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'-
TGTTACCAATCTGAAGTGGGAGCGCCGACCAATTTTTTTTTTTTTTTTTTTT
T-3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 76 a 82 c 133 g 121 t

ORIGIN

Query Match 48.0%; Score 396.6; DB 39; Length 412;
Best Local Similarity 97.8%; Pred. No. 7.2e-100;
Matches 402; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 412 tctcagctaccagctccctccacacgacacagtggttccatgcctaccagttct 471
Db 411 TATCAGCTTACAGACTCCCTCTCCACGGCACAGTGGCTTCCATCTCCTACCACTTC 352

Qy 472 ttgtctatcttcaggaaggaaggtcatctctcttcccaaggaacaaactcgag 531
Db 351 TTGCTTATCTTCCAGGAAGGAAGTCACTCTCTCTCTCCCAAGGAACCAACTCGAG 292

Qy 532 gctctggaataatgacagatttctgaaccgtttccacctggcgacacctgaagcaaga 591
Db 291 GCTCTCGGAATAATGACAGATTCTTGAACCGTTTCCACCTGGCGCAACCTGAAGCAAGCA 232


```
/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pYT73D
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5']-
TGTACCAACTGACGTGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTT
T-3'], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pYT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."
```

```
BASE COUNT 124 a 122 c 109 g 90 t 6 others
ORIGIN

Query Match 46.4%; Score 383.2; DB 26; Length 451;
Best Local Similarity 97.1%; Pred. No. 3.9e-96;
Matches 440; Conservative 0; Mismatches 8; Indels 5; Gaps 5;

Qy 180 attggtcgaagggtgtcttcctgattgtaacaactacagacagagatcacctctcgtggtg 239
Db 2 ATTGGCTGCAAGGTGTTCCTGATGTGAACAACACTACAGACAGAAAGATCACCTCTGGATG 61
Qy 240 gagcgcgatgcaagttcccgggcggtgacgcgcgaacctatctcgtgtaagtgtg 299
Db 62 GAGCGGATAGTCAAGTTCCTCCGGGGCGCTGGACNGC-ANACCTATATCTCGTGATGTG 120
Qy 300 gatccagatgcccttagcagagcagacccagacagagattcttgagacattggcgtgta 359
Db 121 GATCCAGATGCCCTTAGCAGA-NAGAACCCAGACAGATCTTGGAGACATTTGGTGTA 179
Qy 360 acagatatcaaggcgccagacctgaagaagggaagattcaggccagagattatcagcc 419
Db 180 ACAGATATCAAGGGCGCGCCTTGAAGGAGGGAAGATTCAGGGCCAGGAGTTATCAGCC 239
Qy 420 taccaggtccctcccccaccgcgcacagtggtctccatcgctacc-agttcttgctta 478
Db 240 TACCA-GWTCCTCTCCACCGGCACACAGTGGCTTCCATCCGTACCAAGTCTTTGCTA 298
Qy 479 tcttcaggaaggaagggtcatctctctctcccaagaaacaaaactc-gaggctctt 537
Db 299 TCTTCAGGAGGAAAGTCATCTCTCTCTCTCCCAAGGAAACAAACTCAGAGCTCTT 358
Qy 538 ggaataatggacagattctgaacggttccaccctggcggaacctgaagcaagcaccag 597
Db 359 GGAATAATGGACAGATTTCTGAACCGTTTCCACCTGGGGGAACCTGAAGCAAGCAGCC 418
Qy 598 tcatgaccagaactaccagactcaccacccc 630
Db 419 TCATGACCCAGAACTACCAGGACTCANCAACCC 451
```

```
RESULT 12
LOCUS F33888 402 bp mRNA EST 13-MAY-1999
DEFINITION HSPD28217 HM3 Homo sapiens cDNA clone s4000149H10, mRNA sequence.
```

```
ACCESSION F33888
VERSION F33888.1 GI:4819514
KEYWORDS EST.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS Lanfranchi,G., Muraro,T., Caldara,F., Pacchioni,B., Pallavicini,A.,
```

TITLE

JOURNAL
MEDLINE
COMMENT

Pandolfo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.
Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
Genome Res. 6 (1), 35-42 (1996)
On Jun 5, 1998 this sequence version replaced gi:3188423.
Contact: Valle G.
CRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
http://grup.bio.unipd.it.

FEATURES

source

1. 402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s4000149H10"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/note="Vector: pCDNALL (Invitrogen); Site_1: BstXI;
Site_2: NotI; The library was constructed by G.
Lanfranchi. This library is not subtracted nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCGGCTCGAGCGCGCTTTTTTTTTTTTTTTT-3'). The
ds cDNA was sonicated and size-selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adapters. NotI digested and
directionally cloned into BstXI-NotI cut pCDNALL vector."

BASE COUNT 108 a 130 c 86 g 72 t 6 others
ORIGIN

Query Match 46.1%; Score 380.6; DB 50; Length 402;
Best Local Similarity 96.5%; Pred. No. 2e-95;
Matches 386; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 368 caaggcgccagacctgaagaagggaagattcaggggccagagttatcagctaccagcc 427
Db 3 CAAGNNCTTCGACCTGAAGAAACNNAAGATTCAGGACCAGGAGTTATCAGCCTACAGTC 62
Qy 428 tccctcccccacggcacaacagtggttcctcatcgctaccagttcttctctatcttcagga 487
Db 63 TCCCTCCCCACCGGCACACAGTGGCTTCATCGCTACCAAGTCTCTTTGTCTATCTTCAGGA 122
Qy 488 aggaaggtcatctctctctcccaaggaacacaaactcgaggctcttgaaaaatgga 547
Db 123 AGGAAAAGTCATCTCTCTCTCCCAAGAAACAAAACACTCGAGGCTCTTGGAAATGGA 182
Qy 548 cagattctgaacggtttccacctggcggaacctgaagaagcaccagcttcattgaccca 607
Db 183 CAGATTCTGAACCGTTTCCACCTGTGCGAACTGAAGCAAGCACCAGTTCATGACCCA 242
Qy 608 gaactaccagactcaacacccctcaggctcccgagaaagggccagcgcccaagca 667
Db 243 GAATACAGGAGACTCACCACACCTCCAGGCTCCCAAGAAAGGGCCAGCGACCCCAAGCA 302
Qy 668 caaaacccagcgagatagctgcctctagatagcggcttgcctccatccggcgatgtg 727
Db 303 CAAAACCCACGGGAGATAGCTGCTCTAGATAGCCGGCTTTGCCATCCGGGCGATGG 362
Qy 728 ccacactgccaccaccacccagatgtgggtatggaaccccc 767
Db 363 CCACACTGCCACCACCACCGACGATGTGGGTATGGAAACCCCC 402

RESULT 13
F20527
LOCUS F20527
DEFINITION HSPD04880 HM3 Homo sapiens cDNA clone NOTAVAIL04880, mRNA sequence.

EST

13-MAY-1999


```

|||||
Db 280 GGCCTCTTGGAAATGGACAGATTCTGAACCGCTTCCACCTGGCGCAACTGAAGCAAGC 221
|||||
QY 591 acccagttcatgaccagaactaccaggaactaccacacctccaggctccacagagaagg 650
|||||
Db 220 ACCCAGTTCATGACCCAGAACTACCAAGACTACCAACCCCTCCAGGCTCCAGAGGAAGG 161
|||||
QY 651 gccagcgagcccaagcaaaaaccagcgagagatagctgcctgctagatagcgcgcttt 710
|||||
Db 160 GCCAGCAGCCCAAGAC-AAAACCAAGCGGAGATAGCTGCCGTGCTAGATAGCGGGCTTT 102
|||||
QY 711 gccatcgggcatgtggccacatgccacaccagcagcatgtgggtatggaacccctct 770
|||||
Db 101 GCCATCGGGCATGTGGCCACACTGCTCACCACCGCAGCATGTGGGTATGGAACCCCTCT 42
|||||
QY 771 ggatacagaacccctctttttcaataataaaaaa 807
|||||
Db 41 GGATACAGAACCCCTTCTTTTCCAAATTAATAAAAAA 5

RESULT 15
F36903 436 bp mRNA EST 13-MAY-1999
LOCUS HSPD34824 HM3 Homo sapiens cDNA clone SH5-000021-0/A09, mRNA
DEFINITION sequence.
ACCESSION F36903
VERSION F36903.1 GI:4822529
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 436)
Lanfranchi,G., Muraro,T., Caldara,F., Pacchioni,B., Pallavicini,A.,
Pandolfo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.
Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
Genome Res. 6 (1), 35-42 (1996)
96276048
On Mar 10, 1998 this sequence version replaced gi:2948469.
Contact: Valle G.
CRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
http://grup.bio.unipd.it.
FEATURES
source
1..436
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="SH5-000021-0/A09"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/notes="Vector: pcDNAII (Invitrogen); Site:1: BstXI;
Site:2: NotI; The library was constructed by G.
Lanfranchi. This library is not subtracted nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCGGCTCGAGCGCGCTTTTCTTTTCTTTT-3'). The
ds cDNA was sonicated and size-selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adapters, NotI digested and
directionally cloned into BstXI-NotI cut pcDNAII vector."
BASE COUNT 111 a 144 c 84 g 76 t 21 others
ORIGIN

```

Query Match 45.4%; Score 374.6; DB 50; Length 436;
 Best Local Similarity 95.3%; Pred. No. 9.6e-94;
 Matches 383; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

QY 399 caggggccagagttatcagcctaccaggctccctccccaccggcacacagtggcttccat 458
|||||
Db 35 CANNCCCCNAGNTATTTTCTTACCANGCTCCCTCCCAACGGGCAACAGGGGGTCCAT 94
|||||
QY 459 cgctaccagttctttgtctatcttcaggaagaaagtgctatctctctctccctcccaaggaa 518
|||||
Db 95 CGCTACCAGTTCCTTTGTCTATCTTTCAGGAAGGAAAGTCTATTTCTCTCTCCCAAGGAA 154
|||||
QY 519 acaaaactcgaggctcttgaaaatggacagatttctgaacctttccacctggggcga 578
|||||
Db 155 AACAAAACTCGAGGCTCTTGGAAATGGACAGATTTTCTGAACCGTTTCCACCTGGGGA 214
|||||
QY 579 cctgaagcaagcaccagttcatgaccacgaactaccaggactcaccacacctccaggt 638
|||||
Db 215 CCTGAACCAAGCACCCAGTTCATGACCCAGAACTACCAGACTCACCACCTCGAGGCT 274
|||||
QY 639 ccagagaaaaaggggccagcgagcccaagcaaaaaaccaggcgagagatagctgcctgctag 698
|||||
Db 275 CCCAGAGAAAGGGCCAGCGAGCCCAAGCAAAAAACCCAGCGGAGATAGCTGCCTGCTAG 334
|||||
QY 699 atagccggctttgccatcccgggcatgtggccacactgccaccaccacagatgtgggtat 758
|||||
Db 335 ATAGCCGGCTTTGCCATCCGGGCATGTGGCCACACTGCCACCACCACCGACGATGTGGGTAT 394
|||||
QY 759 ggaacccctctggatcacagacccctctttttccaaataaa 800
|||||
Db 395 GGAACCCCTCTTGGATACAGAACCCCTTCTTTTCCAAATAAA 436

```

Search completed: May 23, 2000, 22:04:07
 Job time: 4005 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 21:00:49 ; Search time 6115.43 Seconds
(without alignments)
-277.898 Million cell updates/sec

Title: US-09-215-435-66

Perfect score: 1747

Sequence: 1 gccctaccaatggttcctt.....tggcgcaaaaaaaaaa 1747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_hta1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_hta1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*
51: gb_pr5.*
52: gb_htg8.*
53: gb_htg9.*
54: gb_htg10.*
55: gb_htg11.*
56: gb_htg12.*
57: gb_htg13.*
58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1683	96.3	5051	9	AB032975 Homo sapi
2	1648	94.3	2526	40	AF190725 Homo sapi
3	1648	94.3	3252	51	AF200193 Homo sapi
4	1646.4	94.2	2070	40	AF200343 Homo sapi
5	1646.4	94.2	5878	40	AF201468 Homo sapi
6	1242.6	71.1	2043	12	AF200346 Mus muscu
7	1235.2	70.7	3880	12	AF190726 Mus muscu
8	1141.8	65.4	1506	51	AF204943 Homo sapi
9	1100.6	63.0	2158	12	AF190727 Rattus no
10	973.8	55.7	1114	51	AF161367 Homo sapi
11	739.4	42.3	198092	45	AC020997 Homo sapi
12	737.8	42.2	176874	32	AP000761 Homo sapi
13	735.8	42.1	108772	32	AP000685 Homo sapi
14	725.8	41.5	151455	32	AP000892 Homo sapi
15	297.6	17.0	1804	40	AF200342 Homo sapi
16	297.6	17.0	1873	40	AF117892 Homo sapi
17	296	16.9	1862	51	AF204944 Homo sapi
18	296	16.9	1863	40	AF050171 Homo sapi
19	296	16.9	1885	51	AF200192 Homo sapi
20	296	16.9	2990	51	AF178532 Homo sapi
21	218	12.5	198092	45	AC020997 Homo sapi
22	194.4	11.1	600	13	G24698 human STS W
23	185.2	10.6	972	12	AF051150 Mus muscu
24	77.2	4.4	80580	45	AC021622 Homo sapi
25	77.2	4.4	110000	32	HSMX1.3
26	58.8	3.4	80580	45	AC021622
27	42.4	2.4	1351	12	GP1PROGAS
28	41	2.3	162473	40	AC005803
29	40	2.3	3380	13	G27212
30	40	2.3	4227	9	HUML2TR1
31	39.2	2.2	4501	2	STMX1NA
32	39.2	2.2	38734	1	SC10A5
33	39.2	2.2	125602	33	HS1097F14
34	39	2.2	64418	42	AC015786
35	38.8	2.2	165581	41	AC011448
36	38.6	2.2	268008	44	AC015535
37	38.4	2.2	188800	33	HSJ421D16
38	38.2	2.2	2356	2	AF107888
39	38.2	2.2	36734	1	SC6G10
40	38.2	2.2	68231	55	AC023459
41	37.8	2.2	1276	10	MFPEPC
42	37.4	2.1	171447	41	AC007912
43	37.4	2.1	183217	53	AC012070
44	37.4	2.1	189813	41	AC009410
45	37.2	2.1	184446	42	AC011076

ALIGNMENTS

```

RESULT 1
AB032975
LOCUS Homo sapiens mRNA for KIAA1149 protein, partial cds. 11-NOV-1999
DEFINITION AB032975
ACCESSION AB032975
VERSION AB032975.1 GI:6330044
KEYWORDS
SOURCE
Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:hg01289.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Hirosewa,M., Nagase,T., Ishikawa,K., Kikuno,R., Nomura,N. and
Ohara,O.
TITLE Characterization of cDNA clones selected by the GeneMark analysis
from size-fractionated cDNA libraries from human brain
JOURNAL DNA Res. 6 (5), 329-336 (1999)
MEDLINE 20039618
REFERENCE 2 (bases 1 to 5051)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1999) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/,
Tel:+81-438-52-3913, Fax:+81-438-52-3914)
FEATURES
Location/Qualifiers
1..5051
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hg01289"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
1..1191
/gene="KIAA1149"
<1..1191
/gene="KIAA1149"
/Note="Start codon is not identified."
/codon_start=1
/product="KIAA1149 protein"
/protein_id="BAA8463.1"
/db_xref="GI:6330045"
/translation="YLOAHFTLCGMSSTYRLRGVVPYVYTOGWEGELGPDLSIP
HGNPTVVRNIAITFESDKFFINGSNWEGILGLAYAEIARDDSLPPDLSLVKQTHV
PNLSIQLCGAGPLNQSEVLASVGSMLIGIDHSLSITGSLWYTPIRREYIEVIV
RVEINGQDLKMDCKEYNIDKSIDVSGTNNLRPKVFAAANKSIAASSTEFKPDGFV
LGEQVQWQAGTTPWNIFFVLSYLMGEVNTNSFRITILPOQLRPFVEDVATSDQDCY
KFAISQSSTGTVMGAVIMEGFVYVDFDRARKRIGFAVSACHVHDEFRTAAVEQPFVTL
MEDCGYNIPQTDSTLMTIAYVMAAICALFMLPLCLMWCVQRCLRLRQHDHDFADDI
SLK"
BASE COUNT 1385 a 1207 c 1137 g 1322 t
ORIGIN
Query Match 96.3%; Score 1683; DB 9; Length 5051;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1708; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
QY 25 tatctgaagcccaactttacactctgtcttgggtggtccagcacataccggagacctccgg 84
DB 1 TATCTGCAAGCCCACTTTACACTCTGTTCTGGGTGTTCCAGCACATACCGGACCTCCGG 60
QY 85 aaggggtgtatgtgcctacacccaggcaagtgagggaagggagctgggacccagctg 144
DB 61 AAGGGTGTATGTGGCCCTACACCCAGGCAAGTGCGGAAGGGGAGCTGGGACCCGACCTG 120
QY 145 gtaagcatccccatggcccccaagctcactgtgcgtgccaacattgctcccatcactgaa 204
DB 121 GTAAGCATCCCCATGGCCCCCAACAGTCACTGTGCTGCGGCAACATTGCTGCCATCACTGAA 180

```

```

QY 205 tcagacaagtcttctcatcaacggctccaaactgggaagcatcctggggtggcctatgct 264
DB 181 TCAGACAAGTTCTTCATCAACGGCTCAACTGGNAGGCATCTGGGGCTGGCTATGCT 240
QY 265 gagattgcaaggcctgacgactcccgaggaccttcttttgactctctggtaaagcagacc 324
DB 241 GAGATTGCCAGGCTGACGACTCCCTGGAGCCTTTCTTTGACTCTCTGTGTTAAAGCAGACC 300
QY 325 caggtcccaaacctcttccctgcagcttggagcttggagcttccctcccaacagctct 384
DB 301 CAGTTTCCCAACCTCTCTCCCTGCAGCTTTGTGGTGGCTGGCTTCCCTCCCAACAGTCT 360
QY 385 gaagtctggcctctctgcgaggagcatgatcatctggaggtatcgaccactcgctgtac 444
DB 361 GAAGTGTGCGCTCTCTCGGAGGAGCATGATCATTTGGAGGTATCGACCACTCGCTGTAC 420
QY 445 acaggcagctctctgtatcacacccatcccggggagtggtattatgatgagtgatctgtg 504
DB 421 ACAGGCAGTCTCTGTGTATACACCCATCCGCGGGAGTGTGTTATGAGGTGATCATTTGTG 480
QY 505 cgggtggaatcaatgagcagagatctgaaatgagctgcaagagtgacaaactatgacaag 564
DB 481 CGGTGGAGATCAATGGACAGGATCTGAAATGGACTGCAAGGAGTACAACTATGACAAG 540
QY 565 agcattgtgagcagtgccaccacaccccttctgtttgcccaagaaagtgtttgaagctgca 624
DB 541 AGCATTTGTGACAGTGGCACCACCACCTTCGTTTGGCCCAAGAAAGTGTGTTGAAGTGCA 600
QY 625 gtaaatccatcaaggcagcctcctccacggagaaagtctccctgacggtttcttgcttagga 684
DB 601 GTCAATTCATCAAGCAGCTCTCTCCAGGGAGAAGTTCCTCTGATGTTTCTGGCTAGGA 660
QY 685 gacagctgctgtgctggaagcagcagcaccaccccttggaacattttccagtcactca 744
DB 661 GAGCAGTGTGTGCTGGCAAGCAGGACACCCCTTGGAAACATTTTCCAGTCACTCTCA 720
QY 745 ctctacctaatgggtgaggttaccacaccagctcttcgcgcatcaccatccttcgcgagcaa 804
DB 721 CTCTACCTAATGGTGAGGTACCACACAGTCTCTCCGCATCACCATTCTTCGCGAGCAA 780
QY 805 taactggcgcagtggaagatgtgcccagtcgccacgacgactgttacaagtttggcactc 864
DB 781 TACCTCGGCCAGTGTGAAGATGTGGCAGCTGCCAAGACGACTGTTTACAAGTTTGGCATC 840
QY 865 tcacagtcacacagcagcactgttatggagctgttatcatgagagggcttctacgtgtc 924
DB 841 TCACAGTCAATCCAGGGCACTGTTATGGGAGCTGTTATCATGGAGGGCTTCTACGTTGTC 900
QY 925 ttgtatcgggggcccgaaacgaatttggcttctgtcagcgcgttgccatgtgcacgatgag 984
DB 901 TTTGATCGGGCCCGAAACGAATTTGGCTTTGCTGTGTCAGCGCTTGCCATGTGCACGATGAG 960
QY 985 ttcaggacggcggcgggaagggcccttttgcacctggaacatggaagactgtggcta 1044
DB 961 TTCAGGACGGCAGCGGTGGAAGGCC- TTTTGTGTCACCTTGGACATGGAAGACTGTGGCTA 1019
QY 1045 caacattccacagacagatgagtcacccctcatgacctagcctatgctatgctgctccat 1104
DB 1020 CAACATTCCACAGACAGATGAGTCAACCCCTCATGACCATAGCCTATGTCTATGGCTGCCAT 1079
QY 1105 ctgcccctcttctcatgctccactctgctcatggtgtgtcagtgagcgtgctcctccgtg 1164
DB 1080 CTGGCCCTTCTTCATGCTGCCACTCTGCCTCATGTGTGTCTCAGTGGCGCTGCTCGCTG 1139
QY 1165 cctgcccgcagcagcatgatactttgtctgatgacatctcctcctgctgaagtggaggcccc 1224
DB 1140 CCTGGCCCAAGCAGCATGATGACTTTTGTCTGATGACATCTCCCTGCTCAAGTGAGGAGGCC 1199
QY 1225 atgggcagagatagggtatccccctggaccacacccctcctggttcaacttggatcaact 1284
DB 1200 ATGGGCAGAGATAGAGATTCCCTTGGACACACCTCCCTGGTGGTTTCACTTTGGTCACAAGT 1259

```


QY 360 tgcgtgcttccccctcaaccagctgtaagtgctggtcctctgctcgagggagcatgatcat 419
Db 651 TGCTGGCTTCCCTCCCTCAACCAAGTCTGAAGTCTGCGCTCTCTCGGAGGAGCATGATCAT 710
QY 420 tggaggtatcacacactcgtctgtacacaggcagtcctctgtgtatcacacccatccggcgagga 479
Db 711 TGGAGGTATCAGCACTCGCTGTGTACAGGCGAGTCTCTGTGTATACACCCATCCGCGGGA 770
QY 480 gtggattatgagtgatcattgtcgaggtgagatcaatgagacagatctgaaatgga 539
Db 771 GTGGTATATCAGGTATCATGTGTGCGGTGGAGATCAATGGACAGAGTCTGAAATGGA 830
QY 540 ctgcaaggagtacaactatgacaagagcattgtgagcagtgagccaccacacctctgtt 599
Db 831 CTGCAAGGAGTACAATATGACAAGAGCATTTGTGACAGTGGCACCAACCACTTCGTTT 890
QY 600 gccaaagaagtgttgaagctgcagtcagtcacatccatcaagcagcctctccacggagaa 659
Db 891 GCCCAAGAAAGTGTGGAAGCTGCAGTCAATCCATCCATCAAGGACGCTCTCCACGGAGAA 950
QY 660 gtccctgaaggttctggtctagtagagcagctggtgtctgcaagcagcagcaccacc 719
Db 951 GTTCCCTGATGGTTCTGGCTAGGAGAGCAGCTGTGTGTGCGAAGCAGGACCAACCCC 1010
QY 720 ttgaaacatttcccatctcatctcaactaattggtgaggttaccacacagctcctt 779
Db 1011 TTGGAACATTTTCCACGATCATCTCACTCTACCTAATGGGTGAGTTACCAACCACTCTT 1070
QY 780 ccgcatcacatcttcccgacgaataacctgcggcagtggaagatgtgcccacgtccca 839
Db 1071 CCGCATCACCATCTTCCGACGAATAACCTGCGGCCAGTGGAAAGATGTGCCACGTCCTCA 1130
QY 840 agacgactgttaacagtttgccatctcacagtcacacagcagcagcagctgtatggagctgt 899
Db 1131 AGACGACTGTTACAAGTTCCCATCTCACAGTCAATCCACGGGACACTGTATGGAGCTGT 1190
QY 900 tatcatgagggcttcaactgtctgttctgagtcggccgaaacgaattgcttctgt 959
Db 1191 TATCATGGAGGGCTTCTACCTTCTTGTATCGGGCCGCGAAACGAATGGCTTTGCTGT 1250
QY 960 cagcgtctgcattgcagatagctcagcagcagcagcagcagcagcagcagcagcagcagc 1019
Db 1251 CAGCGCTTGCCATGTGCACGATGAGTTACGAGCGGAGCGGAGGAGGCGCC-TTTTGTCA 1309
QY 1020 ccttgacatggaagactgtggtacacattccacagacagatgagtcacacccatga 1079
Db 1310 CCTTGGACATGGAAGACTGTGGCTACAACATCCACAGACAGATGAGTCAACCCATGA 1369
QY 1080 ccatagcctatgtcagtgccatctgcgcctctctcatgtgccaactctgcctcatgg 1139
Db 1370 CCATAGCCTATGTATGGTGGCCATCTGCGGCCCTTTCATGCTGCCACTGTGCTCATGG 1429
QY 1140 tgtgtcagtgccgtccctccgtcgtcgcagcagcagcagcagcagcagcagcagcagcagc 1199
Db 1430 TGTGTCTAGTGGCGCTCCCTCCGCTGCTGCGCAGCAGCAGATGATGATTTGTGTATGACA 1489
QY 1200 tctcctctgtaagtgagggagccatggtgcagaagatagggattccccctggaccacacc 1259
Db 1490 TCTCCCTGCTGAAGTGAGGAGGCCATGCGCAGAGATAGATATCCCTGTGACCAACACC 1549
QY 1260 tccgtgttcaacttgggtacaaagttaggagacacagatggccactgtggtccagagcaact 1319
Db 1550 TCGTGGTTCATTTGGTCAAGTAGGAGACACAGATGGCCACCTGTGGCCAGAGCACT 1609
QY 1320 caggacccctccccaccacccaaatgcctctgctgtgaggaagaaagagcagcagcagcagc 1379
Db 1610 CAGGACCCCTCCACCACCAATGCCCTCTGCTTGTATGGAGAAAGAAAGGCTGGCAAG 1669
QY 1380 gtgggttcaggagactgtacctgttaggagacaaagagaagaagaagcactctgtg 1439
Db 1670 GTGGGTTCAGGGACTGTACCTGTAGGAAACAGAAAGAAAGAAAGAAAGCACTCTGTGTG 1729

QY 1440 gcgggaataactcttggtcacactcaaatgagtcgggaaattctgctgcttgaacttca 1499
Db 1730 CGGGAATACTCTTGGTCACTCAATTTAAGTCGGGAAATTTCTGCTGCTTGAACCTTCA 1789
QY 1500 gccctgaaccttctgt-caccattcctttaaattctccaaaccccaagattctctcttct 1558
Db 1790 GCCTGAACCTTTGTCCACCATTCCTTTAATTTCTCCACCAAGATTTCTTTCTTCT 1849
QY 1559 tagttcagaagtagctgcatcacacgagcttacccttggcgtgtgctccctggtgaccc 1618
Db 1850 TAGTTTTCAGAAAGTACTGGCATCACACGAGTTACCTTGGCGTGTGTCCCTGTGGTACCC 1909
QY 1619 tggcagaagaagacacacagcttcttccctgctggtgcaaaagtcagtaggagagatgcac 1678
Db 1910 TGCAGAGAAGACCAAGCTTGTTCCTGCTGGCCAAAGTCAGTAGGAGAGATGCAC 1969
QY 1679 agttcctatttctttagacagagcagctgtataaacagcctaacattggtgcacaaa 1738
Db 1970 AGTTTGCTATTTGCTTTAGACAGAGGACTGTATAACAGACCTTACATTTGGTGCAGAA 2029

RESULT 5
AF201468
LOCUS AF201468 5878 bp mRNA PRI 19-DEC-1999
DEFINITION Homo sapiens APP beta-secretase mRNA, complete cds.
ACCESSION AF201468
VERSION AF201468.1 GI:6601444
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 5878)
AUTHORS Sinha, S., Anderson, J. P., Barbour, R., Basl, G. S., Caccavello, R., Davis, D., Doan, M., Dovey, H. F., Frigon, N., Hong, J., Jacobson-Croak, K., Jewett, N., Keim, P., Knops, J., Lieberburg, I., Power, M., Tan, H., Tatsuno, G., Tung, J., Schenk, D., Seubert, P., Sijmensaari, S., Wang, S., Walker, D., Zhao, J., McConlogue, L., and John, V.
TITLE Purification and cloning of amyloid precursor protein
JOURNAL beta-secretase from human brain
MEDLINE Nature 402 (6761), 537-540 (1999)
REFERENCE 2 (bases 1 to 5878)
AUTHORS Basl, G. S., Power, M. D., Wang, S., Tatsuno, G., Frigon, N., Doan, M., Hong, G., Keim, P., Anderson, J., Sinha, S. and McConlogue, L. M.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1999) Gene Expression Group, Elan Pharmaceuticals, Inc., 800 Gateway Blvd., S. San Francisco, CA 94080, USA

FEATURES
source Location/Qualifiers
1..5878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="D11S432"
/tissue_type="brain"
455..1960
/function="cleaves APP at the beta-cleavage site"
/note="membrane type aspartyl protease"
/codon_start=1
/product="APP beta-secretase"
/protein_id="AAF18982.1"
/db_xref="GI:6601444"
/translation="MAQALPWLWMAGVLPVHGTHQIRLPLRSGLGAPLGLRLP
RETDPEPEPRGRGSFVEMVDNLGRKSGQGYVEMTYSPPQTLNLIIVDGSNFAVG
AAPPLHRYQRLSYRDLRLKRVYVPTQKWEGLGTLVSIPIHGNVTVRANI
AATFESDKFTINGSNWEGILGLAYAEIARPDLSLEPFLVQTHVPLFSLQLCGA
GFPLNQSEVLASVGGSMIGGIDHSLYTGSILWTPVIRREAYEYIIVRVEINGDLKM
DKCEYNVDKSIQSDGTTNLRPKVFEAAVKSIAASSTKEKPDGPHGLGQLVCWQAG
TTPNIPFVLSLYLMGEVTVQSPRITILPQYLRPVEDVATSQDCKTFAISOSTGT
VMGAVINEGYVDFDRARRKIRGFAVSACHVHDEFRTAAVGGPFTLDMEDCGNIPQT
DESTLMTIATVMAICALFMLPLCLMLVCMQWRCLRLRQHQHDDFADDDISLTK"

FEATURES
source

Location/Qualifiers

1. .2043

/organism="Mus musculus"

/db_xref="taxon:10090"

CDS

1. .1506

/codon_start=1

/product="aspartyl protease 2"

/protein_id="AAFI7082.1"

/db_xref="GI:6561820"

/translation="MAPALHLLWVSGMLPAQGHGLIRLPLRSLAGPPLGLRLP
REIDEESEPRGRGSFVEMVDNLRKSGCYVYEMTVGSPQPLNLLVDGSSNFAVG
AAPHTLHYRQQLSTYDLRKGYVPTQKWEGLDGLVSLPHGPNVTVRANI
AATSDKPFINSWEGILGLAYAEIARDDSLSEFFDSLVKQTHPIFISQLCGA
GFLPNQTEALASVGGSMIGGIDHSLYTGLMTPIRREMYEVIIVRVEINGQDLKM
DCKEYNDKSIDVSGTTLNLRPKKVFEEAAVKSIIKAASTEKPKPDGFWELVCWQAG
TTPWNPFIISLYLMGEVFNOSPRITLIPQYLRPVEDVATSDDCYKFAVSQSTGT
VMGAVIMEGYVYVDFARKRIGFPAVSACHVHDEFRTAAVEGPFVADMECCYNIPT
DESTLMTIAVMAIICALPMLPLCLMVQWRCLRLRHQHDFFADDDISLLK"

BASE COUNT 453 a 572 c 572 g 446 t
ORIGIN

Query Match 71.1%; Score 1242.6; DB 12; Length 2043;
Best Local Similarity 87.3%; Pred. No. 0;
Matches 1454; Conservative 0; Mismatches 200; Indels 12; Gaps 8;

QY 60 gtccagcacataccgggaccccggaagggtgtgtatgtgccctacacccagggcaagtg 119
DB 351 GTCCAGCACATATCGAGACCTCCGAAAGGCTGTGTATGTCCTTACACCCAGGCAAGTG 410
QY 120 ggaaggaggagctggaccagctgtgaagcatcccccatggccccaaagctactgtg 179
DB 411 GGAGGGGAACCTGGGCAACCGACCTGGTGAGCATCCCTCATGCGCCCAACGTCACCTGTGG 470
QY 180 tgccaaactgtgccaactgaatcagacaagttcttctcatcaacggctcccaactgga 239
DB 471 TGCCAACTATGCTGCCATCACTGAATCGGACAGTCTTCATCATCAATGGTTCACACCTGGGA 530
QY 240 aggcactctgggctggccttatctgagattgcccagcctgagcactccccgagccttt 299
DB 531 GGGCATCTTACGGCTGGCTTATGCTGAGATTGCGAGGCGCGGACGACTCTTTGGAGCCCTT 590
QY 300 ctttgactctctgttaagcagaccacagcttcccacactcttccctcagctttgtgg 359
DB 591 CTTTGTACTCCTGCTGAGCAGACCCACATTCACACATCTTTTCCCTGAGCTCTGTGG 650
QY 360 tgtgtgttccccctcaaccagctgaagtgtgctgctgtgctgagggagcagatgatcat 419
DB 651 CGCTGGCTTCCCTTCAACAGAGGAGGACCTGGGCTCGGTGGGAGGAGCATGATCAT 710
QY 420 tggaggtatcgaccactcgtgtacacagcagctgtctgtgtatatacaccatccggcgga 479
DB 711 TGGTGTATCGACCATCTGCTATACAGGGGAGTCTCTGGTATACACCATCCGCGGGA 770
QY 480 gtggtattatgaggtgatcatgtgctggtgagatcaatgacaggtatctgaatatgga 539
DB 771 GTGGTATTATGAGTATGATTTGACGTGTGAATCAATGCTCAAGATCTCAAGATGGA 830
QY 540 ctgcaaggagtgtaactatgacagagcatgtggacagtggaaccacacacttcgttt 599
DB 831 CTGCAAGGAGTACAATAGCAGCAAGAGCATTTGGACAGTGGGACCAACCTTCGCTT 890
QY 600 gcccaagaaagtgtttgaagctgagctcaaatccatcaaggcagcctctccacagagaa 659
DB 891 GCCCAAGAAAGTATTGTGAAGCTGCGCTCAAGTCCATCAAGGAGGAGCTCTCGACGAGAA 950
QY 660 gttcctgacggtttctggtgagagagcagctggtgtgctggtggaagcagcaccacccc 719
DB 951 GTTCCCGATGCTTTTGGCTAGGGGAGCAGCTGTGTGCTGGCAAGCAGCAGCAGCCCC 1010
QY 720 ttggaacattttccagctcatctactactactaatggtggtgaggttaccacacagctctt 779
DB 1011 TTGGAACATTTTCCAGCTATTCTACCTTACCTCATGTGGTGAAGTCAACCAATCAGTCTT 1070

QY 780 ccgcatcacatcccttccgcagacaataccctgcggccagtggaagatgtggccacgtccca 839
DB 1071 CCGCATCACCATCTCTCTCAGCAATACCTACGCGCGGTGGAGCAGCTGCCACGTCCCA 1130
QY 840 agacgactgttaacaagtgttccatctcacagtcacacggtcagcggtgactgttatgagctgt 899
DB 1131 AGAGGACTGTTACAGTTCGCTGTCTCACAGTATCCACGGGCACTGTTATGGAGGCGGT 1190
QY 900 tataatgaggggttcttaoactgttcttggatcgggccggccgaaacgaattgctttgctgt 959
DB 1191 CATCATGGAAGGTTCTATGCTCTTCGATCGAGCCGGAAGCAATTTGGCTTTGCTGT 1250
QY 960 cagccttgcactgtgcacagatagttccaggaaggcagcggtggaagggccnnttttgtca 1019
DB 1251 CACGCTTGCCATGTGCACAGATGAGTCAAGGACGGGCGGCAAGAGGTCCG-TTTGTTA 1309
QY 1020 ctttgacatgggaagactgtggctacaacattccacagacagatgagtcacacccatca 1079
DB 1310 CGGCAGACATGGAAGACTGTGGCTACACATTTCCACAGACATGAGTCAACACTTATGA 1369
QY 1080 ccatgcttatgtcatgggtgcoactgtgcctcttctcatgtgccactctgctcctatgg 1139
DB 1370 CCATAGCCTATGTATCGGGGCGCATCTGCGCCCTCTTCATGTTGCCACTCTGCTCATGS 1429
QY 1140 tgtgtcagtgggcgtcctccgtcgtcgtcgcagcagcagcatgatgactttgtgtgaca 1199
DB 1430 TATGTCAGTGGGCTGCTGCGGTGCTTGGCTGGCCACACCATGACTTTTGTGTGATGACA 1489
QY 1200 tctccctgctgaagtggagggcccatgggcagagaagtagggattcccttgagcacacc 1259
DB 1490 TCTCCTGCTCAAGTAAGGAGGCTCGTGGGAGATGATGGAGTGTGATGCTGTGGCAGAGCAC 1549
QY 1260 tccgtgttcaactttgtgtcaca--agtaggagacagatggcaactgtggtgcagagcac 1317
DB 1550 TGGGTGGTTCCCTTTGGTGCATATGAGTTGGAGCTATGAGTGTGATGCTGTGGCAGAGCAC 1609
QY 1318 ctcaagaccctcccccacaccccaaatgctcgtcctgtatgaggaagaaaggtggca 1377
DB 1610 CTCAGGACCTCACCAACCTGCGCAATGCTTCTGGGTGACAGA-ACAGAGAAATCAGGCA 1668
QY 1378 aggtgggttccaggactgtacctgtagagacagaaaaagaaagaaagacactctgc 1437
DB 1669 AGCTGGATTACAGGCTTGCACTGTAGCACACAGAGAGGGAAGAGCAGCGTCTGG 1728
QY 1438 tggcggtgaactctgtgtcactcaaatgctggtggtggtggtggtggtggtggtggtggt 1497
DB 1729 TGGCAGGAATATCTTAGGCACCACTTGAGT-TGGAAATTTTGTGCTTGAAGCTT 1787
QY 1498 cagcctgaaacctttgtcaccattctttaaattctcaaccacaaagtattctctcttctt 1557
DB 1788 CAGCCTCAGCCTCTGCGCCAGCATCTTTAGAGTCTCCAACTAAAGTATTCTTTATGTC 1847
QY 1558 ttatttcagaagactgtgcatcacacagcaggttaccttggctgtgtgtgtgtgtgtgtgtgt 1617
DB 1848 ---CTTCCAGAGTACTGGCGTCACTACTCAGGCTACC-CGGCATGTGTCTCCTGTGTACC 1903
QY 1618 ctgcaagagaagacacaaagtgtttccctgctggccaaagttagtaggagagatgca 1677
DB 1904 CTGCAGAGAAGGGCCAAATC-TCAATTCCTGCTGGCCAAAGTCAAGAGAAAGAGGTG-- 1961
QY 1678 cagttgtctatttcttttagagacagggactgtataaacaaagccta 1723
DB 1961 AAGTTTGCCAGTTGCTTTAGTGATAGGACTGACACTCAAGCCTA 2006

RESULT 7

AF190726

LOCUS

DEFINITION

cds

ACCESSION

VERSION

AF190726 3880 bp mRNA

Mus musculus beta-site App cleaving enzyme (Bace) mRNA, complete

AF190726

AF190726.2 GI:6760476

27-JAN-2000

ROD

enzyme (Bace)

mRNA, complete

KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3880)
AUTHORS Vassar,B.D., Bennett,B.D., Babu-Khan,S., Kahn,S., Mendiaz,E.A.,
Denis,P., Teplow,D.B., Ross,S., Amarante,P., Loeloff,R., Luo,Y.,
Fisher,S., Fuller,J., Edenson,S., Lile,J., Jarosinski,M.A.,
Biere,A.L., Curran,E., Burgess,T., Louis,J.-C., Collins,F.,
Teanor,J., Rogers,G. and Citron,M.
TITLE Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
the transmembrane aspartic protease BACE
JOURNAL Science 286 (5440), 735-741 (1999)
MEDLINE 2002972
REFERENCE 2 (bases 1 to 3880)
AUTHORS Bennett,B.D., Vassar,R. and Citron,M.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1999) Neuroscience, Amgen Inc., One Amgen Center
Dr., Thousand Oaks, CA 91320-1799, USA
REFERENCE 3 (bases 1 to 3880)
AUTHORS Bennett,B.D., Vassar,R. and Citron,M.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2000) Neuroscience, Amgen Inc., One Amgen Center
Dr., Thousand Oaks, CA 91320-1799, USA
REMARK Sequence update by submitter
COMMENT On Jan 27, 2000 this sequence version replaced gi:6118540.
FEATURES
source
1..3880
/organism="Mus musculus"
/db_xref="taxon:10090"
gene
1..3880
/gene="Bace"
CDS
428..1933
/gene="Bace"
/codon_start=1
/product="beta-site APP cleaving enzyme"
/protein_id="AAF04143.2"
/db_xref="GI:6760477"
/translation="MAPALHLLWVSGMLPAGQTHLGLRLPSRLGAPPLGLRLP
REIDSEERGRGSFVEMVDNLKRGVYVEMTVGSPPTNLILVDLTVGSSNFVAG
AAPHFLLHYRQLSTYDLRKGVTYPTQKWEGLGDLVSLPHGNVYVRANI
AAITESDKFLINGSNWGIIGLAIETARPDDSLPEFFDSLVKTHIPNIFSLQCGA
GFPLNORFALASGMSIIGIDHSLVTGSLWTPIRREWYEVIIIVRVEINGDLKM
DCKEYNKSDIVSGTNNLRPKKVFEEAVKSAASSTKFPDGFGLVQLVQWQAG
TTPWNIPVLSLVMGTNQSFRITILPQOYLRPVEDVATSDQCKFPAVSQSTGT
VMGAVIMEGFIYVDFARKRIGFAVSACHVHDEFRTAAVSGPFVTDMECDGYNIPOT
DESTLMTIAYVMAICALFMLPLCLMYCQWRCLRLRHQHDDEFADDSILK"
BASE COUNT 893 a 1083 c 1022 g 882 t
ORIGIN
Query Match 70.7%; Score 1235.2; DB 12; Length 3880;
Best Local Similarity 86.5%; Pred. No. 0;
Matches 1456; Conservative 0; Mismatches 214; Indels 13; Gaps 8;
Qy 60 gtccagacataccggagcctccggaagggtgtgtatgtgcctacacccagggcaagt 119
Db 778 gtccagacacatatccagacctcccaagggtgtgtatgtgcctacacccagggcaagt 837
Qy 120 ggaagggaagctgggacccagctggtaagcatcccccatggcccccaagctgactgtgcg 179
Db 838 GGAGGGGAAGTGGGCACCGACCTGGTGAGGATCCCTCATGCGCCCAACGTCATGTGCG 897
Qy 180 tgcacaactgtgctgcatcactgaatcagacaagttcttcacacgagctcccaactggga 239
Db 898 TGCCAACATGCTGCCATCACTGAATCGACAAAGTCTTCATCAATGTTCCAACTGGGA 957
Qy 240 aggaatcctgggctgctgtatgtgagattgccaggcctgacgactccccggagccttt 299
Db 958 GGGATCTCTAGGGTGGGCTATGCTGAGATTGCCAGGCCGACGACTCTTTGGAGCCCTT 1017
Qy 300 ctttgactctctgtaagacagacccacgcttcccaaccttctccctgcagcttttgg 359

Db 1018 CTTTGACTCCCTGGTGAAGCAGACCCACATTCCTCAACATCTTTTCCCTGCAGCTCTGTGG 1077
Qy 360 tgcgtgtctccccctcaaccagctggaagtgctgcctctctgcggaggagcatatcat 419
Db 1078 CGCTGGCTTCCCTCCCAACACAGACCCAGTTCGCTCGGTGGAGGAGCATGATCAT 1137
Qy 420 tggaggtatccaccactcgctgtacacaggaagtcctcgtgtatatacaccatccgcgga 479
Db 1138 TGGTGTATCGACCACTCGCTATACCGGGCAGTCTCTGGTACACCCATCCGGGGGA 1197
Qy 480 gtggtattatgaggtgcatattgtcggggtgagatcaatggacaggatctctgaaatgga 539
Db 1198 GTGGTATTATGAAGTGATCAATTGTACGTGTGAAATCAATGGTCAAGATCTCAAGATGA 1257
Qy 540 ctgcaaggagtacaactatgacaagagcatgtggaagtggaagtggaagtggaagtgga 599
Db 1258 CTGCAAGGAGTACAACACTACGACAAAGAGCATTTGTGACAGTGGGACCAACCTTCGCTT 1317
Qy 600 gcccaagaagtgtttgaagctgcagtcgaatcccaatcccaagcagcctcctccacgagaa 659
Db 1318 GCCCAAGAAAGTATTTGAAGCTGCCGTCAAGTCCATCAAGGACGAGCTCTCCACGGAGAA 1377
Qy 660 gtccctgacggtttctctgtaggagagcagctggtgtgtgctgcaagcaggcaccacccc 719
Db 1378 GTTCCCGGATGGCTTTGGTAGGGAGCAGCTGCTGCTGCAAGCAGGACGACGCC 1437
Qy 720 ttgaaacattttccagctcatctcactaataatgggtgaggttacaacacagctcctt 779
Db 1438 TTGGAACATTTTCCAGTCAATTTTCACTTTTACCTCATGGTGAAGTCAACCAATCAGTCTT 1497
Qy 780 ccgcatcacctctccgcagacaatactgcgcagcaatactgcgcagcaatactgcgcagca 839
Db 1498 CCGCATCAACATCTTCTCTCAGCAATACCTACGCGCGGTGGAGGACGTGGCCACGTCCTCA 1557
Qy 840 agacgactgttacaagtttgccatctcagctcagctcctccacggcagctgttatggagctgt 899
Db 1558 AGACGACTGTTACAAGTTCGCTGTCTCAGCTATCCACGGGCACTGTTATGGAGCGGT 1617
Qy 900 tateatgagggctctcagctgtgtgttgatggcgccgcaaaaacgaattggtttgctgt 959
Db 1618 CATCATGGAAGGTTTCTATGTCTTCGATCGAGCCCGCAAAACGAAATTTGGCTTGTCTGT 1677
Qy 960 cagccttgccatgtgcagcatgagttcagacgcagcgggtggaagcccttttggca 1019
Db 1678 CAGCCTTGCATGTGCAGATGAGTTCAGGACGGCGGCGGAGGAGTCCG-TTGTGTTA 1736
Qy 1020 ccttgacatggaagctgtggctacaactccacagacagatgagtcacacccctcatga 1079
Db 1737 CGGCAGACATGGAAGACTGTGGCTACACATTTCCCGACAGACATGAGTCAACACTTATGA 1796
Qy 1080 caatagcctatgtcattggtgcctatctgcgcctcttctcatgtgcacctctgcctcattgg 1139
Db 1797 CCATAGCTATGTATGCGCGCATCTGCGCCTCTTTCATGTTGCCACTCTGCTCATGG 1856
Qy 1140 tgtgtcagtggtgcctcctgcctcctgcgcagcagcatgactgttctgtgatgaca 1199
Db 1857 TATGTGAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1916
Qy 1200 tctccctctgaagtgaaggagggcccatggcagaagatagggattccccctggaccacacc 1259
Db 1917 TCTCCCTGCTCAAGTAAGAGGCGCGTGGGACAGATGATGAGAGCGCCCTGACACATC 1976
Qy 1260 tccgtggttactcttgggtaca--agtaggagacacagatggcactgtggtggcagagcac 1317
Db 1977 TGGGTGGTTCCTTTGGTTCACATGAGTTGGAGCTATGGATGTTACTGTGGCCAGAGCAC 2036
Qy 1318 ctcaaggacccctccccaccccaataatgctcctctgtccttgatggagagaaagagctggca 1377
Db 2037 CTCAGGACCTCACCAACCTGCAATGCTTCTGCGGTGTCAGAACAGAGAAAT-CAGGCA 2095
Qy 1378 aggtgggttccaggagctgtacctgtagagacagaaagagaaagagacactctgc 1437


```

QY 1020 ccttggacatggaagactgtggctacacattccacagacagatgagtcacaccctcatga 1079
Db 1310 CTTTGGACATGGAAGACTGTGGCTACAAATTCACACAGACAGATGAGTCAACCCCTCATGA 1369
QY 1080 ccatagctatgtatggctgccaatctgcccctctctcatgctgccaatctgctcatagg 1139
Db 1370 CCAATAGCTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1429
QY 1140 tgtgtcagtggtgctgctcctgctgctgctgctgctgctgctgctgctgctgctgctgct 1199
Db 1430 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1489
QY 1200 tctccctgctgaaatga 1216
Db 1490 TCTCCCTGCTGAAGTGA 1506

RESULT 9
AF190727 LOCUS AF190727 2158 bp mRNA ROD 26-OCT-1999
DEFINITION Rattus norvegicus beta-site APP cleaving enzyme (Bace) mRNA,
complete cds.
ACCESSION AF190727
VERSION AF190727.1 GI:6118542
KEYWORDS .
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 2158)
Vassar, P., Bennett, B.D., Babu-Khan, S., Kahn, S., Mendiaz, E.A.,
Denis, P., Teplow, D.B., Ross, S., Amarante, P., Loeloff, R., Luo, Y.,
Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M.A.,
Biere, A.L., Curran, E., Burgess, T., Louis, J.-C., Collins, F.,
Freednor, J., Rogers, G., and Citron, M.
Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
the transmembrane aspartic protease BACE
Science 286 (5440), 735-741 (1999)
2002972
2 (bases 1 to 2158)
Bennett, B.D., Vassar, R. and Citron, M.
Direct Submission
Submitted (29-SEP-1999) Neurosciences, Amgen Inc., One Amgen Center
Dr., Thousand Oaks, CA 91320-1799, USA
JOURNAL Location/Qualifiers
FEATURES
source 1..2158
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
gene 1..2158
/gene="Bace"
CDS 428..1933
/gene="Bace"
/codon_start=1
/product="beta-site APP cleaving enzyme"
/protein_id="AAF04144.1"
/db_xref="GI:6118543"
/translation="MAPALEWLLWVSGMLPAQCTHLGIRLPLRSLAGPPGLRLP
REFDEPPEPRGRGSFVNDNRKSGGQYIVEMTVGSPQFLNILDVIGSSNFVAG
AAPHPFLRYQRLSTYNDLRKSVYPTQGWEGELGTLDSIPHGNVTVRANI
AAITSDKFFINGSNWEGILGLAYAEIARPDLSLEFFDSLVKQTHIPNIFSLQCGA
GFPLNQTALASVGGSMIIGGIDHSLYTGSLWYTPIRREWEYEVIVRVEINGDQLKM
DCXEYNVDSIVSGTNNLRPKKVFEEAAVKSJKAASSTKPKPDGFGLRQLVCMWAG
TTPWNIFPISLVMGEVINOQSFRIPLPOQLRPVEDVATSDDCYKFKVASSSVGT
VMGAVMEGIVYLVDFARKRIGFAVSACHVHDFEFAAVEGPFVTFADMEDCGNIPQT
DESTLMTIAYVMAIICALFMPLCLMVCWMQRCLRHQHDFFADDISLLK"
BASE COUNT 458 a 640 c 619 g 441 t
ORIGIN

Query Match 63.08; Score 1100.6; DB 12; Length 2158;
Best Local Similarity 88.2%; Pred. No. 1.5e-280;
Matches 1219; Conservative 0; Mismatches 160; Indels 3; Gaps 2;

```

```

QY 60 gtccagcacatacccggaacccctccggaagggtgtgtatgtgcccctacaccccgagcgcaatg 119
Db 778 GTCCAGTACATACCGAGACCTCCGAAAGTCTGTATGTGCCCCTACACCCAGGGCAAGTG 837
QY 120 ggaagggaagctggggcaccgacctggtaagcatcccccatggcccccaacgctcaactgtgcg 179
Db 838 GGAGGGGAACATGGGCACTGACCTGGTGAGCATCCCTCATGGCCCCAACGCTCACTGTGCG 897
QY 180 tgccaaatctgctcccatcactgaatcagacaaagtctcttcacatcaggctccaactggga 239
Db 898 TGCCAAACATTTCTGCCATCACTGAATCGGAAGTCTTTCATCAATGTTTCCAACTGGGA 957
QY 240 aggcattcctggggctggcctatgtgagattgccaggcctgacgactcccccggagccttt 299
Db 958 GGGCATCCTAGGGCTGGCCTATGCTGAGATTGCCAGGCTGACGACTCTTTGGAGCCCTT 1017
QY 300 ctttgactctctgtaaagcagacccacgcttcccaacctctctccctgcagcttgtg 359
Db 1018 TTTTGAAGTCTCCCTGGTGAAGCAGACCCACATTCGGAACATCTTTTCCCTGCAGCTGTG 1077
QY 360 tgcctggcttccccctcaaccagctctgaagtctgagctgctgctgctgagggagcagatcat 419
Db 1078 CGCTGGCTTCCCCTCAACCAGACTGAGGCACTGGCCCTCGGTGGAGGGAGCATGATCAT 1137
QY 420 tggagggtatcgaccactcgctgtacacaggcagctctctctgtgtatcacaccatccggcgga 479
Db 1138 TGGTGGTATCGACCATTTCCCTATACACTGGCAGTCTCTGGTACACACCCATCCGCGGGA 1197
QY 480 gtgtattatgagtgatcattgtgcgggtgagatcaatggacaggaatcctgaataatgga 539
Db 1198 GTGCTATTATGAAGTGATCAATTTACGTGTAGAAATCAATGCTCAAGATCTGAAATGGA 1257
QY 540 ctgcaaggagtaacactatgacagagcattgtgagacagtgaggacacacacacctcgctt 599
Db 1258 CTGCAAGGAGTACAACATATGACAAGAGCATCGTGACAGTGGCAGCCCTCTCTGACGGGAA 1317
QY 600 gcccaagaaagtgttgaagctgagtcacaaatccatcaagcagcagcctctccacggagaa 659
Db 1318 GCCCAAGNAAGTATTGNAAGCTGCAGTCAAGTCCATCAAGCAGCGCTCTCTGACGGGAA 1377
QY 660 gtccctgacgggttcttgctgtaggagcagctggtgtgctggcaagcagcagcaccccc 719
Db 1378 GTTCCCGGATGGCTTTTGGCTAGGGAGCAGCTGGTGTGCTGGCAAGCAGCAGCAGCCCC 1437
QY 720 ttgaaacatttccagctcatctcaactcatcctaataatgggtgaggttaccacaccagctct 779
Db 1438 TTGAAACATTTTCCAGTCAATTTTCACTTTTCACTCATGCGGTGAAGTCAACCAATCATGCT 1497
QY 780 ccgcatcacatcctccgacgaataacctgcccagctggaagatgtggcagctccca 839
Db 1498 CCGCATCACCATCTCTCTCCAGCAATACCTACGGCCAGTGGGAAGATGTGGCCACGTCCTCA 1557
QY 840 agacgactgttacaaagtgtgcatctcactcatccacggcgactgttatggagctgt 899
Db 1558 AGACGACTGTTTACAAGTTTCGCGCTCTCAGCTCATCCACAGCAGCCGTTATGGAGCGGT 1617
QY 900 tatcatgagggcttctacgtgtgtctttgtatgagggccgcaaaacgaatgtgctttgtcgt 959
Db 1618 CATCATGGAAGCGTCTTATGTGGTCTTTTGTATGAGCCCGGAAAGCAATTTGGCTTTGCTGT 1677
QY 960 cagccttgccatgtgacagatgagttcagacagcagcagcagcagcagcagcagcagcagcag 1019
Db 1678 CAGCGCTTGCCATGTGACAGATGAGTTTCAGGACGCGGAGTGGGAAGTTCGG -TTTGCA 1736
QY 1020 ccttggacatggaagactgtgctacaaatccacacagacagatgagtcacaccctcatga 1079
Db 1737 CGGCAGACATGGAAGACTGTGGCTACAAATTCACACAGACAGATGAGTCAACACTTATGA 1796
QY 1080 ccatagcctatgtcatggctgccaatctgcccctctctctctctctctctctctctctctctct 1139
Db 1797 CCATAGCCTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1856
QY 1140 tgtgtcagtggtgctgctcctgctgctgctgctgctgctgctgctgctgctgctgctgct 1199

```



```

RESULT 11
AC020997/c
LOCUS AC020997 198092 bp DNA HTG 25-JAN-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-677N11 map 11, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
AC020997
ACCESSION AC020997.2 GI:6751743
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 198092)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,Y., Collangeio,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehotzky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 25, 2000 this sequence version replaced gi:6691361.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5120
Center clone name: 677_N_11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190312 bases at least Q40
Consensus quality: 195004 bases at least Q30
Consensus quality: 196725 bases at least Q20
Insert size: 209000; agarose-fp
Insert size: 198092; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1213: contig of 1213 bp in length
*
*

```

```

* 1214 3303: contig of 2090 bp in length
* gap of unknown length
* 3304 5848: contig of 2545 bp in length
* gap of unknown length
* 5849 8316: contig of 2468 bp in length
* gap of unknown length
* 8317 11798: contig of 3482 bp in length
* gap of unknown length
* 11799 15125: contig of 3327 bp in length
* gap of unknown length
* 15126 19072: contig of 3947 bp in length
* gap of unknown length
* 19073 22877: contig of 3805 bp in length
* gap of unknown length
* 22878 26589: contig of 3712 bp in length
* gap of unknown length
* 26590 33219: contig of 6630 bp in length
* gap of unknown length
* 33220 39283: contig of 6064 bp in length
* gap of unknown length
* 39284 49390: contig of 10107 bp in length
* gap of unknown length
* 49391 59154: contig of 9764 bp in length
* gap of unknown length
* 59155 71025: contig of 11871 bp in length
* gap of unknown length
* 71026 82461: contig of 11436 bp in length
* gap of unknown length
* 82462 99502: contig of 17041 bp in length
* gap of unknown length
* 99503 121381: contig of 21879 bp in length
* gap of unknown length
* 121382 153661: contig of 32280 bp in length
* gap of unknown length
* 153662 198092: contig of 44431 bp in length.
FEATURES
Location/Qualifiers
1. 198092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-677N11"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 52207 a 46145 c 48779 g 50961 t
ORIGIN
Query Match 42.3%; Score 739.4; DB 45; Length 198092;
Best Local Similarity 99.0%; Pred. No. 8.4e-185;
Matches 765; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
Qy 967 tggcatgtgcacgatgagttcaggacgagcggtgggaagccenttttgcaccttggga 1026
Db 95075 TCCCAGTGCACGATGAGTTTCAGGACGCGGTTGAAGGCC-TTTTGTACCTTGA 95017
Qy 1027 catggaagactgtggctacacattccacagacagatgagtcacccctcatgaccatagc 1086
Db 95016 CATGGAAGACTGTGGCTACAACATTCCACAGACAGATGAGTCAACCCCTCATGACCATAGC 94957
Qy 1087 ctatgtcatggtgcctatgcgcctcttcatgctgcacctctgcctcatggtgtgca 1146
Db 94956 CTATGTCTATGCTGCCATCTCGGCCCTCTTTCATGCTGCCACTCTGCCCTCATGGTGTGTCA 94897
Qy 1147 gtggcgctccctccgctccctgcgcgcagcagcatgatgacttttgcctgatgacatccct 1206
Db 94896 GTGGCGCTGCCCTCGCTGCCCTGCCGCGCAGCAGCATGATGACTTTGTGCTGATGACATCTCCT 94837
Qy 1207 gctgaagtgaaggagcccatgggacagaagataggatctccctcgaccacacacctccgtgg 1266
Db 94836 GCTGAAGTGAAGGAGGCCCATGGCAGAGATAGATATCCCTCGGACACACACTCCGTGG 94777
Qy 1267 ttccatttggtcacaagttagagacacagatggcacctgtggcagagacacctcaggacc 1326

```

Db 94776 TTCACCTTTGGTCACAAAGTAGGAGACACAGATGGCACCCTGTGSCCAGAGACACCTCAGGACC 94717
QY 1327 CTCcccccccccacaaatgcctctgcttgatgagagaagaaagctgcaaggtgggtt 1386
Db 94716 CTCcccccccccacaaatgcctctgcttgatgagagaagaaagctgcaaggtgggtt 94657
QY 1387 ccaggactgtacctgtaggagacagaaagagaaagaaagcaactctgctgcgggaa 1446
Db 94656 CCAGGACTGTACCTGTAGGAACAGAAAGAGAAAGAACACTCTGCTGCGGGAA 94597
QY 1447 tactcttgctcacctcaaatgaagtcgggaaattctgtctgtctgcttgaacttcagccctga 1506
Db 94596 TACTCTTGGTCACCTCAAAATTAAGTCGGGAAATTCGTCTTGAACCTTCAGCCCTGA 94537
QY 1507 accttgt-caccattccttaattctcaaccacaaagtattcttcttcttagtttc 1565
Db 94536 ACCTTTGTCCACCAATCTTTAAATTCCTCAACCAAGTAATCTCTTTCTTTAGTTTC 94477
QY 1566 agaagtactggcatcacacgcaggttacctggcgtgtgtccctgtgtaccctggcaga 1625
Db 94476 AGAAGTACTGTCATCACAGGCAGGTTACCTTGGCGTGTGCCCTGTGTCACCTGGCAGA 94417
QY 1626 gaagagacaaagctgttctcctctgctgcccacaaagtcagtaggagagatgcacagtttc 1685
Db 94416 GAAGAGACCAAGCTGTGTTCCTCTGCTGCGCAAGTCACTAGTAGGAGGATGCACAGTTTC 94357
QY 1686 tatttctttagacagaggactataaacaagcctaaactggtgcacaaa 1738
Db 94356 TATTGTCTTAGACAGACGGGACTGTATAACAGCCCTTAACATTGGTGCAAGA 94304

RESULT 12
AP000761/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-794I11 map 11q23, WORKING
DRAFT SEQUENCE, 41 unordered pieces.
ACCESSION
AP000761.1 GI:6997599
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens DNA, clone:RP11-794I11.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 176874)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Homo sapiens 176,874 genomic DNA of 11q23
Published Only in Database (1999) In press
2 (bases 1 to 176874)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Direct Submission
Submitted (25-NOV-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Feb 18, 2000 this sequence version replaced gi:6469067.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@sc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-794I11
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 131265 bases at least Q40
Consensus quality: 145088 bases at least Q30

Consensus quality: 153043 bases at least Q20
Insert size: 157890; sum-of-contigs
Quality coverage: 4.00x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 41 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 17060 contig of 17060 of in length
17561 26431 contig of 8871 of in length
26932 37174 contig of 10243 of in length
37675 50315 contig of 12641 of in length
50816 59520 contig of 8705 of in length
66674 66674 contig of 6654 of in length
67175 72749 contig of 5575 of in length
73250 79383 contig of 6134 of in length
79884 85508 contig of 5625 of in length
86009 91395 contig of 5387 of in length
91896 95164 contig of 3269 of in length
95665 99111 contig of 3447 of in length
99612 103924 contig of 4323 of in length
104435 108668 contig of 4234 of in length
109169 112967 contig of 3799 of in length
113468 116560 contig of 3093 of in length
117061 120239 contig of 3179 of in length
120740 123392 contig of 2653 of in length
123893 127142 contig of 3250 of in length
127643 130250 contig of 2608 of in length
130751 132561 contig of 1811 of in length
133062 135186 contig of 2125 of in length
135687 137591 contig of 1905 of in length
138092 139516 contig of 1425 of in length
140017 141828 contig of 1812 of in length
142329 144664 contig of 2336 of in length
145165 147389 contig of 2225 of in length
147890 149432 contig of 1543 of in length
149933 151748 contig of 1816 of in length
152249 154238 contig of 1990 of in length
154739 156951 contig of 2213 of in length
157452 158677 contig of 1226 of in length
159178 161160 contig of 1983 of in length
161661 163453 contig of 1793 of in length
163954 165709 contig of 1756 of in length
166210 167786 contig of 1577 of in length
168287 169682 contig of 1396 of in length
170183 171628 contig of 1446 of in length
172129 173191 contig of 1063 of in length
173692 175227 contig of 1536 of in length
175728 176874 contig of 1147 of in length.

* NOTE: This is a 'working draft' sequence. It currently consists of 41 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 17052: contig of 17052 bp in length
* 17053 17560: gap of 508 bp
* 17561 26392: contig of 8832 bp in length
* 26393 26931: gap of 539 bp
* 26932 37168: contig of 10237 bp in length
* 37169 37677: gap of 509 bp
* 37678 50309: contig of 12632 bp in length
* 50310 50815: gap of 506 bp
* 50816 59520: contig of 8705 bp in length
* 59521 60021: gap of 501 bp
* 60022 66674: contig of 6653 bp in length
* 66675 67175: gap of 501 bp
* 67176 72738: contig of 5563 bp in length

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 108772)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens 108,772 genomic DNA of 11q23
 Published Only in Database (1999) In press
 2 (bases 1 to 108772)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (08-NOV-1999) to the DDBJ/EMBL/GenBank databases.
 Masahira Hattori, The Institute of Physical and Chemical Research
 (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail: hattori@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-42-778-9923, Fax: 81-42-778-9924)
 On Feb 18, 2000 this sequence version replaced gi:6277513.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@sc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11
 Center clone name: CMB9-8M6
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 83833 bases at least Q40
 Consensus quality: 90336 bases at least Q30
 Consensus quality: 93974 bases at least Q20
 Insert size: 93875; sum-of-contigs
 Quality coverage: 4.26x in Q20 bases; sum-of-contigs

COMMENT

NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and the order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 10546 contig of 10546 bp in length
11047 17615 contig of 5569 bp in length
18116 24086 contig of 5971 bp in length
24587 29736 contig of 5150 bp in length
30237 38079 contig of 7843 bp in length
38580 43515 contig of 4936 bp in length
44016 48745 contig of 4730 bp in length
49246 53614 contig of 4369 bp in length
54115 59119 contig of 5005 bp in length
59620 63763 contig of 4144 bp in length
64264 71186 contig of 2923 bp in length
67687 71093 contig of 3407 bp in length
71594 74598 contig of 3005 bp in length
75099 78391 contig of 3293 bp in length
78892 81382 contig of 2491 bp in length
81883 83560 contig of 1678 bp in length
84061 86317 contig of 2257 bp in length
86818 88559 contig of 1742 bp in length
89060 91079 contig of 2020 bp in length
91580 94007 contig of 2428 bp in length
94508 96680 contig of 2173 bp in length
97181 99237 contig of 2057 bp in length
99738 101029 contig of 1292 bp in length
101530 102776 contig of 1247 bp in length
103277 104568 contig of 1292 bp in length
105069 107111 contig of 2043 bp in length
107612 108772 contig of 1161 bp in length.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 10544: contig of 10544 bp in length
10545 11048: gap of 504 bp
11049 17615: contig of 5567 bp in length
17616 18115: gap of 500 bp
18116 24086: contig of 5971 bp in length
24087 24587: gap of 501 bp
24588 29734: contig of 5147 bp in length
29735 30236: gap of 502 bp
30237 38079: contig of 7843 bp in length
38080 38580: gap of 501 bp
38581 43514: contig of 4934 bp in length
43515 44015: gap of 501 bp
44016 48740: contig of 4725 bp in length
48741 49245: gap of 505 bp
49246 53597: contig of 4352 bp in length
53598 54114: gap of 517 bp
54115 59118: contig of 5004 bp in length
59119 59622: gap of 504 bp
59623 63763: contig of 4141 bp in length
63764 64267: gap of 504 bp
64268 67186: contig of 2919 bp in length
67187 67686: gap of 500 bp
67687 71088: contig of 3402 bp in length
71089 71593: gap of 505 bp
71594 74598: contig of 3005 bp in length
74599 75100: gap of 502 bp
75101 78386: contig of 3286 bp in length
78387 78900: gap of 514 bp
78901 81382: contig of 2482 bp in length
81383 81883: gap of 501 bp
81884 83552: contig of 1669 bp in length
83553 84078: gap of 526 bp
84079 86317: contig of 2239 bp in length
86318 86818: gap of 501 bp
86819 88552: contig of 1734 bp in length
88553 89059: gap of 507 bp
89060 91079: contig of 2020 bp in length
91080 91591: gap of 512 bp
91592 94007: contig of 2416 bp in length
94008 94507: gap of 500 bp
94508 96680: contig of 2173 bp in length
96681 97181: gap of 501 bp
97182 99237: contig of 2056 bp in length
99238 99737: gap of 500 bp
99738 101029: contig of 1292 bp in length
101030 101529: gap of 500 bp
101530 102774: contig of 1245 bp in length
102775 103276: gap of 502 bp
103277 104566: contig of 1290 bp in length
104567 105068: gap of 502 bp
105069 107111: contig of 2043 bp in length
107112 107614: gap of 503 bp
107615 108772: contig of 1158 bp in length.

```

FEATURES

Location/Qualifiers
 1..108772
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="CMB9-8M6"
 /map="11q23"

BASE COUNT 25040 a 21557 c 20852 g 26525 t 14798 others
 ORIGIN

Query Match 42.1% Score 735.8; DB 32; Length 108772;
 Best Local Similarity 98.6% Pred. No. 7.1e-184;
 Matches 762; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY	967	tgccatgtgcacgtgagttcagacgagcagcggtggaagccncttttgcaccttggg	1026
Db	51209	TCCCCAGTCACGATCAGTTCAGGACGCGGTGGAGGCC-TTTTCTCACCTTGG	51151
QY	1027	catggaagactgtggtacacacattccacagacagatgagtcacacccctcatgaccatagc	1086
Db	51150	CATGGAAGACTGTGGTACAACTTCACAGACAGATGATCAACCTCATGACCATAGC	51091
QY	1087	ctatgtcatggtgcacatgcgcctcttcctgctgcacactgcctcatggtgtgca	1146
Db	51090	CTATGTCATGGCTGCCATCTGCGCCCTCTTTCATGCTGCCACTCTGCTCATGCTGTGTC	51031
QY	1147	gtggcgctgcctcgtcgtcgcgcagcagcatgatgactttgtgtgacatctccct	1206
Db	51030	GTGGTCTGCCCTCCGTGCTGCGCCAGCAGCATGATGACTTTGCTGATGACATCTCCCT	50971
QY	1207	gctgaagtgaggagggcccatggtggcagaagatagggattccctggaccacacccctcgtg	1266
Db	50970	GCTGAAGTGAGGAGGCCCATGGCAGAAATAGAGATTCCCTTGGACCACCATCTCGTGG	50911
QY	1267	ttaactttggtcaagttaggagacacagatggccactgtgcccagacacacccatcagacc	1326
Db	50910	TTACACTTGTGTCAAGTAGGAGACACAGATGGCACCCTGTGGCCAGAGCACCTCAGGACC	50851
QY	1327	ctccccaccacccaaatgcctctgccttgcgtgagaggaagagctggcaaggtgggt	1386
Db	50850	CTCCCCACCACCAATGCTCTGCCTTGATGGAGGAAGAGCTGGCAAGGTGGGTT	50791
QY	1387	ccaggagactgtacctgttaggagacagaaaaagaaagaaagaaagaaagaaagaaagaa	1446
Db	50790	CCAGGAGACTGTACCTGTAGGAAACAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAG	50731
QY	1447	tactctgtgtcaactcaaaatttaagtcgggaaattctgctgtgaaacttcagccctga	1506
Db	50730	TACTCTGTGTGCACCTCAAAATTAAGTCGGGNAATCTGCTGTGAACATTCAGCCCTGA	50671
QY	1507	accttgtt-caccatctcttaattctccaaacccaaagtattctctcttcttagtttc	1565
Db	50670	ACCTTTGTCCACCATCTCTTAATTTCTCAACCCAAAGTATTCTTCTTTCTTAGTTTC	50611
QY	1566	agaagtactggcatcacacgcaggttaccttgcctggcgtgtgtcccttgggtacccttgcaga	1625
Db	50610	AGAAGTACTGGCATCACACGAGTTACCTTGGCTGTGTGTCCTGTGGTACCTCGGCAGA	50551
QY	1626	gaagagaccaagcttcttccctgcctggccaaagtcaagtcagtaggagagatcacagatttc	1685
Db	50550	GAAGAGACCAGCTGTGTTCCCTGCTGGCCAAAGTCAGTAGGAGAGATGCACAGTTTC	50491
QY	1686	tattgttttagagacagggactgtataaacagccctaacatttggtgcaaaa	1738
Db	50490	TATTGTCTTTAGAGACAGGGACTGTATAACACAGCCCTAACATTGGTGCAAGA	50438
RESULT	14		
LOCUS	AP000892	151455 bp DNA HTG 04-FEB-2000	
DEFINITION	Homo sapiens chromosome 11 clone CMB9-94B1 map 11q23, WORKING DRAFT SEQUENCE, 28 unordered pieces.		
ACCESSION	AP000892		
VERSION	AP000892.1	GI:6997716	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens DNA, clone:CMB9-94B1.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 151455) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Homo sapiens 151,455 genomic DNA of 11q23		
JOURNAL	Published Only in DataBase (1999) In press		
REFERENCE	2 (bases 1 to 151455) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (17-DEC-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel.81-42-778-9923, Fax:81-42-778-9924)
On Feb 18, 2000 this sequence version replaced gi:6630630.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: CMB9-94B1
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 121163 bases at least Q40
Consensus quality: 130696 bases at least Q30
Consensus quality: 135650 bases at least Q20
Insert size: 138512; sum-of-contigs
Quality coverage: 4.29x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved
1 12522 contig of 12522 of in length
13023 27386 contig of 14364 of in length
27887 42306 contig of 14420 of in length
48042 contig of 5236 of in length
56254 contig of 7712 of in length
63106 contig of 6352 of in length
69329 contig of 5723 of in length
75476 contig of 5847 of in length
81796 contig of 5820 of in length
86568 contig of 4272 of in length
91893 contig of 4825 of in length
97236 contig of 4843 of in length
102308 contig of 4772 of in length
107140 contig of 4132 of in length
110263 contig of 2623 of in length
113914 contig of 3151 of in length
117714 contig of 3300 of in length
122319 contig of 4105 of in length
127011 contig of 4192 of in length
131419 contig of 3908 of in length
134746 contig of 2827 of in length
137635 contig of 2389 of in length
140554 contig of 2419 of in length
142988 contig of 1934 of in length
145289 contig of 1801 of in length
147398 contig of 1609 of in length
147899 149515 contig of 1617 of in length
150016 151455 contig of 1440 of in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 12507: contig of 12507 bp in length
* 12508 13022: gap of 515 bp
* 13023 27386: contig of 14364 bp in length
* 27387 27895: gap of 509 bp

CDS	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="21"				
	1..1557				
	/codon_start=1				
	/product="aspartyl protease 1"				
	/protein_id="AAFI7078.1"				
	/db_xref="GI:6561812"				
	/translation="MGALARALLPLLAQWLRAAPETPLRVAATNRVVA				
	PTPCPTPAPERHADGLALEPALASPAGNAFLAMVDNLQDGSGRYYLEMLIGTPP				
	QKLIILVDGSSNFVAGTPHSYIDYFDTERSSTYRSKGFDTVYKYTQGSWTGFGE				
	DLVTIPKGTENFLNIATIFESNEFLPIKRWNGILGLAYATLAKPSLSLETFPDSL				
	VTOANIPFMSQMGAGLIPVAGSNGSLVGLGIEPSLYKGDWYITPKIEWYIQI				
	EILKLEIGQSLNLDREYNADKAIIVDSGTLRLRPQKVFDAVVEAVARASLPEFSD				
	GFWTGSOLACWNSETPWSVFPKISYILRDENSSRSFRITILPOLYIQPMGAGLANE				
	CYRFGISPTNALVIGATWEGFVIFDRQKRVGAAPCAEIAAGAAVEISGFST				
	EDVASNCVPAQSLSEFILMIIVSVALMSVCGAILLVLLVLLLPFCRRPRDPEVVND				
	ESSLVRHRWK"				
	BASE COUNT	397 a	520 c	458 g	429 t
	ORIGIN				
	Query Match 17.0%; Score 297.6; DB 40; Length 1804;				
	Best Local Similarity 59.5%; Pred.No. 4.5e-68;				
	Matches 545; Conservative 0; Mismatches 359; Indels 12; Gaps 2;				
	QY	53	ctgggtggtccagacacacgcgggaccccggaagggtgtgtatgtgcccacacccagg	112	
	Db	395	CAGAGAGGCTTAGCACATACCGCTCCAAGGCGTTTGACGTGCACAGTGAAGTACACACAAG	454	
QY	113	gcaagtgggaaggagctgggcacgcacctggtaagcatcccccatggccccaaacgtca	172		
Db	455	GAACCTGGACGGGCTTCGTTGGGGAACACCTCGTCAACCATCCCCAAAGGCTTCAATACTT	514		
QY	173	ctgtgcgtgccaaactctgtgcacatcactgaatcagacaagtcttccatcaaaaggtcca	232		
Db	515	CTTTTCTTGTCACATTCGCCACTATTTTGAATCAGAGAAATTTCTTTTGCTGGGATTA	574		
QY	233	actggaagcactcctgggtggcctatctgagatgcccagcctgacgacctccccgg	292		
Db	575	AATGGAATGGAATACTTGGGCTAGCTTATGCCACACTTGGCAAGCCATCAAGTTCTCTGG	634		
QY	293	agcctttcttgactctctgttaaagcagacccacgttcccaacctctctccctgcagc	352		
Db	635	AGACCTTCTTCGACTCCCTGGTGACACAAGCAACATCCCCAACGTTTCTCCATGCAGA	694		
QY	353	tttgtgtgctggttccccctcaacagctctgaagtctgtggcctctgtcggaggagca	412		
Db	695	TGTGTGGAGCGGCTTGCCGT-----TGCTGGATCTGGGACCAACGGAGGTAGTC	745		
QY	413	tgatcatggaggtatcgaccactcgtgtacacaggcagctctgtgtatcacacccatcc	472		
Db	746	TTGTCTTGGGTGGAAATGGAACCAAGTTTGTAATAAGGAGAGACATCTGGTATACCCCTATTA	805		
QY	473	ggcgggagtggtattagtggtgatcattgtgcgggtggagatcaatggacaggaactga	532		
Db	806	AGGAAGAGTGGTACTACCAGATAGAAATCTCGAAATTTGGAATTTGGAGGCCAAAGCCTTA	865		
QY	533	aaatggactgcaaggaggtacaactatgacaagagcattgtggacagtggcaccaccaacc	592		
Db	866	ATCTGGAGCTGCAGAGAGTATAACGCAGACAAGGCCATCTGGACAGTGGCACCACGCTGC	925		
QY	593	ttcgtttgcccaagaagtgtttgaagctcagtcagtcacatccatcaaggcagcctcccca	652		
Db	926	TGCGCTGCCCAAGAGGTGTGTATCGGTGTGGAGGTGTGGCCCGCGCATCTCTGA	985		
QY	653	cggagaagtctccctgcagcgtttctcgtgtaggagagcagctggtgtgtgcgaagcagga	712		
Db	986	TTCCAGAAATCTCTGATGGTTCGGAAGTCTGGAGTCCCGAGCTGGCGTGTGGACGAATTCGG	1045		
QY	713	ccacccttggaacattttcccaggtcactctacctaataatgggtgaggtttaccacac	772		
Db	1046	AAACACCTTGTGCTTACTTCCCTAAAATCTCCATCTACCTGAGAGATGAGAACTCCAGCA	1105		

Qy	773	agtccttccgcataccacatccttccgcagcaataacctcgcgcagtggaagatgtggcca	832
Db	1106	GGTCATTCCGTATCACAAATCCTGCTCAGCTTTACATTCAGCCCATG---ATGGGGGCGG	1162
Qy	833	cgccccagacgactgtttacaagtttgccatctcacagtcacccacgsggcactgttatgg	892
Db	1163	GCCTGAATTATGAATGTTACCGATTTCGGCATTTCCCCATCCACAAATCGCGCTGGTGATCG	1222
Qy	893	gagctgttatcatagagggcttctacgttctgttgcgccccgcgaaaaacgaattggct	952
Db	1223	GTGCCACGGTATGGAGGCTTCTACGTCATCTTCGACAGAGCCCCAGAAAGGGGTGGGCT	1282
Qy	953	ttgctgtcagcgcttg	968
Db	1283	TCGCAGCGAGGCCCTCG	1298

Search completed: May 24, 2000, 00:02:22
Job time: 10893 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 21:01:54 ; Search time 282.38 seconds
(without alignments)
1547.863 Million cell updates/sec

Title: US-09-215-435-66
Perfect score: 1747
Sequence: 1 gctcaccaatgttccctt.....tggtgcaaaaaaaaaaaaaa 1747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1629	93.2	2541	1 V41696	Nucleotide sequenc
2	1611.4	92.2	2370	1 V41697	Partial nucleotide
3	296	16.9	1862	1 V27962	Aspartic proteinase
4	165.4	9.5	185	1 X39493	Human secreted pro
5	38.2	2.2	1506	1 T64547	NADP-specific glut
6	36.6	2.1	1473	1 T64548	NADP-specific glut
7	36.6	2.1	1969	1 T64531	Glutamate dehydrog
8	36.6	2.1	2096	1 T64543	NADP-specific glut
9	36.6	2.1	2099	1 T64530	NADP-specific glut
10	36.6	2.1	2137	1 T64542	NADP-specific glut
11	36.6	2.1	2140	1 T64529	NADP-specific glut
12	35.6	2.0	2288	1 V59498	Human sodium depen
13	34.8	2.0	12585	1 V62392	Human interleukin-
14	34	1.9	2734	1 Q39221	Encodes human epid
15	34	1.9	2734	1 Q47376	hETG mRNA. Diagnos
16	33.8	1.9	1284	1 V30457	Chimeric HLA-A2.1/
17	33.6	1.9	12912	1 T11708	Polycystic kidney
18	33.6	1.9	14060	1 T94102	Human PKD1 cDNA. H
19	33.6	1.9	14148	1 T13821	Polycystic kidney
20	33.6	1.9	53526	1 T94101	Human PKD1 gene. H
21	33.6	1.9	53577	1 T18551	Human polycystic k
22	33.6	1.9	53577	1 T94108	Human PKD1 locus b
23	33.4	1.9	2151	1 V23269	Human latheo prote
24	33.4	1.9	2152	1 V23271	Human latheo prote
25	33	1.9	405	1 V15495	Aeromonas caviae p
26	33	1.9	1598	1 Q40461	E.coli mdh promote
27	33	1.9	3187	1 V15499	Aeromonas caviae p
28	33	1.9	3187	1 V15498	Aeromonas caviae p
29	32.6	1.9	1128	1 V71135	dnaN gene encoding
30	32.4	1.9	2790	1 V18480	BOPI cDNA. New iso
31	32.4	1.9	3923	1 Q03463	cDNA encoding N-ac
32	32.2	1.8	12036	1 Q04668	FHA structural gen
33	32	1.8	1450	1 V18820	Human hyaluronidas
34	32	1.8	1956	1 V02997	Murine Ena-VASP 11

RESULT 1
V41696
ID V41696 standard; cDNA; 2541 BP.
AC V41696;
DT 26-OCN-1998 (first entry)
DE Nucleotide sequence of human ASP2 (aspartic protease 2).
KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
KW prohormone processing; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1506
FT /*tag= a /product= "human ASP2"
FT EP-855444-A2.
PD 29-JUL-1998.
PF 27-JAN-1998; 300573.
PR 28-JAN-1997; GB-001684.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Chapman CG, Murphy K, Powell DJ, Smith TS;
DR WPI: 98-389809/34.
DR P-PSDB: W59807.
PT New nucleic acid encoding human aspartic protease 2 - used to treat,
PT prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
PT processing
PS Claim 2; Pages 6-7; 26pp; English.
CC This is the nucleotide sequence of the human ASP2 (aspartic protease 2),
CC used in the method of the invention. Agonists and antagonists for
CC ASP2 immunospecific antibodies are used to treat conditions requiring
CC increased or decreased activity or expression of ASP2 respectively.
CC ASP2 is used to treat and diagnose e.g. Alzheimer's disease, cancer
CC and prohormone processing and ASP2 or a fragment can be used to induce
CC an immune response against the above conditions.
SQ Sequence 2541 BP; 598 A; 673 C; 675 G; 579 T;

Query Match 93.2%; Score 1629; DB 1; Length 2541;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1668; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
Qy 60 gtccagacataccgggacccctccggagggtgtgtatgtgcctacacccagggaagt 119
Db 351 GTCCAGCACATACCGGGACCTCCGGAGGGTGTGTATGAGCCCTACACCCAGGGCAAGTG 410
Qy 120 ggaaggagctgggacccgacctggtaagcatcccccattggcccaacgctcactgtgcg 179
Db 411 GGAAGGGAGCTGGGACCGGACCTGGTGAAGCATCCCCCATGGCCCAACGTCACGTGTGCG 470
Qy 180 tgccaacattgtgcctcactgaatcagacagaagttcttcacacgggtcccaactggga 239
Db 471 TGCCAACATTGCTGCCATCACTGATCAGACAGTCTTCATCAGCGCTCCAACCTGGGA 530
Qy 240 aggcattcctgggctgacctatgtgtagattgccaggcctgacgactccccggagccttt 299
Db 531 AGGCATCTGGGGTGGGCTATGCTGAGATTGCCAGGCTGACGACTCCCTGGGACCTTT 590

Human EHOC-1 cDNA.
Enterococcus faeca
Expressed Sequence
Human brain Expres
Desulfurococcus ph
Human bone formati
Human bone morphog
Immunoglobulin D-r
Human steroid-21-h
Human clavipes draglin
N.clavipes draglin

QY 300 ctttgactctctggttaagcagaccacccagtcgttcccaacctcttctccctgcagctttgtg 359
Db 591 CTTTGACTCTCTGGTAAAGCAGACCCAGCTTCCCAACCTCTCTCCCTGCAGCTTGTGG 650
QY 360 tgcctggctccctccaccagctcgaagctcgtgcctctctcggagggaagcagatgacat 419
Db 651 TGCTGGCTTCCCTTCAACCAAGCTCTGAAGTGTGGGCTCTCTCGGAGGGAGCATGATCAT 710
QY 420 tggaggtatcacaccactcgtgtacacaggcagtcctctcgtgtatacacaccatccggcgga 479
Db 711 TGGAGGTATCGACCACTCGCTGTACACAGCGAGTCTCTGGTATACACCCATCCGGGGGA 770
QY 480 gtggtattatgagtgatcatgtgcgggtgagatcaatggacagatctgaatatgga 539
Db 771 GTGGTATTATGAGTGTATCATGTGTGGGGTGGAGATCAATGGACAGGATCTGAAATGGA 830
QY 540 ctgcaagagatcacactatgacaagcagctgtgacagtgagcagtgccaccacacctctgtt 599
Db 831 CTGCAAGGAGTACAACATATGACAAGAGCATTTGTGGCAGTGGCAGCCACCAACCTTCGTTT 890
QY 600 gcccaagaaagtgtttgaagctgcagtcataatccatcaaggcagcctcctccacggagaa 659
Db 891 GCCCAAGAAAGTGTGAAGCTGCAGTCAAAATCCATCAAGGCGACCTCTCCAGCGAGAA 950
QY 660 gtccctgacggtttctgctagagagcagctggtgtcgtgcgaagcagcgaccacccc 719
Db 951 GTTCCCTGATGGTTCTGGCTAGGAGAGCAGCTGTGTGTGGCAAGCGCACACCC 1010
QY 720 ttggaacattttcccaagtcatctcaactcacttaagtgtgaggttaccaaccagtcctt 779
Db 1011 TTGGAACATTTTCCCACTCATCTCACTTACCTTAATGGGTGAGTTTACCAACCAAGTCCTT 1070
QY 780 ccgcatcacactcctccgcagcaataacctgcgcagtggaagatgtgcccacgtccca 839
Db 1071 CCGCATCACCATCTTCCGAGCAATACCTTGCGCCAGTGAAGATGTGSCCAGCTCCCA 1130
QY 840 agacgactgtacaagtttccatctcaacagtcacagtcacagggcagctgttatggagctgt 899
Db 1131 AGAGGACTGTACAAAGTTTGCCATCTCACAGTCAATCCACGGGCACGTATTATGGAGCTGT 1190
QY 900 taccatggagggtctctacgttctcttgatcgccgcccgaacgaatggcttctgct 959
Db 1191 TATCATGGAGGGTCTTACGTTGCTTTGATCGGCCCGCCGAAACGAATTTGGCTTGTCTGT 1250
QY 960 cagcgttgcatgtgcacagatagttcaggaagcagcggtggaagcccttttgtca 1019
Db 1251 CAGCGCTTGCATGTGCACGATGATTCAGGAGCGGAGCGGTGGAAGGCC -TTTTGTCA 1309
QY 1020 ccttgacatggaagactgtggtacaacattccacagacagatgagtaaccctcatga 1079
Db 1310 CCTTGGACATGGAAAGACTGTGGCTACAACTCCACAGACAGATGAGTCAACCCCTCATGA 1369
QY 1080 ccatgcctatgtcatgtgctgcatctgcgcctcttcatgctgccactctgcctcatgg 1139
Db 1370 CCATAGCCATGTATCATGGTGGCATCTGCGCCCTCTTCATGCTGCCACTGCCCTCATGG 1429
QY 1140 tfgtcaagtggcgtcctccgctgctcgcgcagcagcatgatgactttgtcatgaca 1199
Db 1430 TGTGTGAGTGGGCTGCCCTCCGCTGCCAGCAGCATGATGACTTTGTGTGATGACA 1489
QY 1200 tctccctgctgaagttagagggccatgggcagaagataggtattccctt-ggaccacac 1258
Db 1490 TCTCCCTGCTGAAGTGAGGAGGCCCATGGGACAAGATAGATTTCCCTTGGGACACAC 1549
QY 1259 ctccgtggttcaacttgggtcaagtaggagacacagatggcacctgtggccagagcacc 1318
Db 1550 CTCCGCTGTTTCACTTTTGTGTCAAAAGTAGGAGACACAGATGGCACCTGTGGCCAGACACC 1609
QY 1319 tcagacacctccccaccacacaaatgcctctccttgatggagaaggaaggtggcaa 1378
Db 1610 TCAGGACCTTCCCCACCCACCAATGCTCTGCTCTGTATGGAGAGGAAAGGCTGGCAA 1669
QY 1379 gdtgggttccaggagactgtacctgtaggagacagaaaagaaagaaagcactctgct 1438

Db 1670 GGTGGGTTCCAGGGACGTACCTGTAGAAACAGAAAAGAGAAAGACACTCTGTCT 1729
QY 1439 ggcgggaataactcttggtcacacctcaaaatttaagtcgggaaattctgcttctgaaacttc 1498
Db 1730 GCGGGGAATACTCTTGTGTACCTCAAAATTTAAGTCGGGAATTTCTGCTGTTGAAACTTC 1789
QY 1499 agccctgaacacctttgt-cacattcctttaaatctccaaacccaaagtattctcttttc 1557
Db 1790 AGCCCTGAACCTTTGTCCACCATTCTCTTAATTTCTCAACCCAAAGTATCTCTCTTTC 1849
QY 1558 ttagtttcagaagtaactggtcagcagcaggttaccttgctggtgctgctccttggtacc 1617
Db 1850 TTAGTTTTCAGAAGTACTTGGCATCACGCGAGTACCTTGGCGTGTCTCCTGTGGTACC 1909
QY 1618 ctggcagagagaagacaagcttgttccctgctgcccacaaagtcagtaggagagatgca 1677
Db 1910 CGGGCAGAGAAGAGAGCAAGCTTGTTCCTCTGCTGCCAAAGTCAGTAGGAGAGATGCA 1969
QY 1678 cagtttgctatttcttagagacagggactgtataaaacaaagcctaacttggtgcacaa 1737
Db 1970 CAGTTTGCTATTGCTTTTAGAGACAGGACTGTATAACAAGCCTAACATTGGTGCAAG 2029
QY 1738 a 1738
Db 2030 A 2030
RESULT 2
V41697
ID V41697 standard; cDNA; 2370 BP.
AC V41697;
DT 26-OCT-1998 (first entry)
DE Partial nucleotide sequence of human ASP2 (aspartic protease 2).
KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
KW prohormone processing; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 1..2370
FT CDS
FT /*tag= a
FT /product= "partial ASP2 protein"
FT /transl_except= (pos:2284..2286,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2290..2292,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2308..2310,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2326..2328,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2335..2337,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2338..2340,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2341..2343,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2344..2346,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2350..2352,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2353..2355,aa:Xaa)
FT /note= "Xaa= unknown"
FT /transl_except= (pos:2368..2370,aa:Xaa)
FT /note= "Xaa= unknown"
FT misc_feature 1333..1335
FT /*tag= b
FT /note= "stop codon specified in the protein"
FT 1474..1476
FT /*tag= c
FT /note= "stop codon specified in the protein"
FT misc_feature 1598..1599
FT /*tag= d
FT /note= "stop codon specified in the protein"

```
FT misc_feature 1624..1626
FT /*tag= e
FT /note= "stop codon specified in the protein"
FT misc_feature 1816..1818
FT /*tag= f
FT /note= "stop codon specified in the protein"
FT misc_feature 1843..1845
FT /*tag= g
FT /note= "stop codon specified in the protein"
FT misc_feature 1873..1875
FT /*tag= h
FT /note= "stop codon specified in the protein"
FT misc_feature 1891..1893
FT /*tag= i
FT /note= "stop codon specified in the protein"
FT misc_feature 2005..2007
FT /*tag= j
FT /note= "stop codon specified in the protein"
FT misc_feature 2011..2013
FT /*tag= k
FT /note= "stop codon specified in the protein"
FT misc_feature 2029..2031
FT /*tag= l
FT /note= "stop codon specified in the protein"
FT misc_feature 2152..2154
FT /*tag= m
FT /note= "stop codon specified in the protein"
FT misc_feature 2164..2166
FT /*tag= n
FT /note= "stop codon specified in the protein"
FT misc_feature 2179..2181
FT /*tag= o
FT /note= "stop codon specified in the protein"
FT misc_feature 2234..2236
FT /*tag= p
FT /note= "stop codon specified in the protein"
FT misc_feature 2332..2334
FT /*tag= q
FT /note= "stop codon specified in the protein"
PN EP-855444-A2.
PD 29-JUL-1998.
PF 27-JAN-1998; 300573.
PR 28-JAN-1997; GB-001684.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Chapman CG, Murphy K, Powell DJ, Smith TS;
DR WPI; 98-389809/34.
DR P-ESDB; W59808.
PT New nucleic acid encoding human aspartic protease 2 - used to treat,
PT prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
PT processing.
PS Disclosure; Pages 8-9; 26pp; English.
CC This is the nucleotide sequence of the partial human ASP2 (aspartic
CC protease 2), used in the method of the invention. Agonists and
CC antagonists for ASP2 immunospecific antibodies are used to treat
CC conditions requiring increased or decreased activity or expression
CC of ASP2 respectively. ASP2 is used to treat and diagnose e.g.
CC Alzheimer's disease, cancer and prohormone processing and ASP2 or a
CC fragment can be used to induce an immune response against the above
CC conditions.
SQ Sequence 2370 BP; 579 A; 605 C; 609 G; 561 T;
```

Query Match 92.2%; Score 1611.4; DB 1; Length 2370;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1657; Conservative 0; Mismatches 21; Indels 3; Gaps 3;

```
Qy 60 gtcagacacacccggagccctccgaaaggtgtgtatgtgccctacacccagggaagtg 119
Dy 180 gtccagacacataccgggaccctccgaaaggtgtgtatgtgccctacacccagggaagtg 239
Qy 120 ggaagggagctgggacacccgacctggttaagcatcccccatggcccccaactgctgctg 179
```

```
Db 240 GGAAGGGGAGCTGGGCAACCGACCTGGTAAGCATCCCCATGGCCCCCAACGTCACCTGTGCG 299
Qy 180 tgcaaacattgctgccatcactgaatcagacaagttcttcacacggtcacaactgga 239
Db 300 TGCCAACATTGCTGCCATCACTGAATCAGACAAGTCTTCATCAACGGCTCCAACCTGGGA 359
Qy 240 aggcacacctggggctgctatgctgagattgccaggcctgacgactccccgagccctt 299
Db 360 AGGCATCTGGGGCTGGCTATGCTGAGATTGCCAGGCTGACGACTCCCTGGAGCCCTT 419
Qy 300 cttgactctctgtataagcagaccacgcttcccaacctctctccctgacgtcttg 359
Db 420 CTTTGACTCTCTGTGTAAGACAGACCCACGCTTCCCAACCTCTTCCCTGCAGCTTTGG 479
Qy 360 tgcggcttccccctcaacagctgaaagtctgactgctctgctcgaggagacatcat 419
Db 480 TGCTGGCTTCCCTTCAACAGTCTGAAGTGTGCTGCCCTCTGTGCGAGGAGCATGATCAT 539
Qy 420 tggaggtatcgaccactcgctgtacacaggcagctctctggtatatacaccatccggcgga 479
Db 540 TGGAGGTATCGACCACTCGCTGTACACAGGCAGTCTCTGGTATACACCATCCGCGGGA 599
Qy 480 gtggtattatgaggtgatcatgtgctggtgagagatcaatggacaggatctgaaatgga 539
Db 600 GTGGTATTATGAGGTGATCATTTGCGGGTGGAGATCAATGGACAGGATCTGAAATGGA 659
Qy 540 ctgcaaggagtacaactatgacaagacgattgtgacagtgaggcaccacaaacttcgtt 599
Db 660 CTGCAAGGAGTACAACATATGACAAGAGCATTTGTGACAGTGGCACACACCACTTCGTTT 719
Qy 600 gcccaagaaagtgtttgaagctgcagtcacaaatcccatcaaggcagctcctccacggaga 659
Db 720 GCCCAAGAAAGTGTGTAAGCTGCACTCAATCCATCAAGCGAGCTCTCCACGGGAGNA 779
Qy 660 gtccctgacggttctctgctaggagagacgctgtgtctgctggcaagcaggcaccacccc 719
Db 780 GTTCCCTGATGTTTCTGGCTAGGAGAGCAGCTGCTGTGCTGGCAAGCAGGACCACCCC 839
Qy 720 ttggaacatttcccagtcactcactacataatgggtgaggttaccacaaacttcctt 779
Db 840 TTGGAACATTTTCCAGTCATCTCCTACTACCTAATGGGTGAGTTACCAACAGTCTCT 899
Qy 780 cgcataccactcctccgcagcaataactcgcgcagtcagtggaagatgtggcacctccca 839
Db 900 CCGCATCACCATCTCTCCGACCAATACCTGGGGCCAGTGGGAAGATGTGGCCACCTCCCA 959
Qy 840 agacgactgttacaagtttgccatctcacagtcacatccacgggcactgttatgggagctgt 899
Db 960 AGACGACTGTTACAAGTTTGCCATCTCAGTCTACCTCCACGGGCACCTGTTATGGAGCTGT 1019
Qy 900 tatcatggagggtctctacgctgttcttgatggggcccgaaaaaactgtgtcttgcgt 959
Db 1020 TATCATGGAGGGCTTCTACGTTGCTTTGATCGGGCCCGAAAACGAATTTGGCTTTGCTGT 1079
Qy 960 cagccttgccatgtgcacgattgctcagacgacgagcgtgtggaagcccttttgcata 1019
Db 1080 CAGCCCTTGCCATGTGCACGATGAGTTTCAGGACGCGCGGTGGGAAGGCCCT-TTTTGTCA 1138
Qy 1020 ccttgacatggaagactgtggctacaacattccacacagacagatgagtcacacccatga 1079
Db 1139 CCTTGGACATGGAAGACTGTGGCTACAACATTTCCACACAGACAGATGAGTCAACCCCTCATGA 1198
Qy 1080 caatgacctatgtcatgctgacctctgcgcctcttcatgtgcactctgcctcatg 1139
Db 1199 CCATAGCCTATGTCTATGCTGTCATCTGCGCCCTCTTATGCTGTCACCTCTCCCTCATGG 1258
Qy 1140 tgtgtcagtgccgtgcctccgctacctgcgcacgacgacgacgacgacgacgacgacgac 1199
Db 1259 TGTGTGAGTGGCGCTGCTCCGCTGCTGCGCCACGACAAATGATGACTTTGCTGTGATGACA 1318
Qy 1200 tctccctgctgaagtgaaggagggcccatggcagaagatagggattccccct-ggacacac 1258
Db 1319 TCTCCCTGCTGAAGTGAAGGAGGCCCATGGGAAAGATAGAGATTCCCTTGGGACCACAC 1378
```

QY 1259 ctccggtgttcacttttgggtacaaagtagagacacagatggcacctgtggccagagcacc 1318
 |||||
 Db 1379 CTCCTGGTTCATCTTGGTTCACAAGTAGGACACAGATGGCACCTGTGGCCAGAGCACC 1438
 |||||
 QY 1319 tcagagacctcccccaccacaaatgctctgcttgatggagaagaaagggctggcaa 1378
 |||||
 Db 1439 TCAGGACCTCCACCACCACCAATGCTCTGCTTGATGGAGAGGAAAGGCTGGCAA 1498
 |||||
 QY 1379 gttggttccagggactgtacctgtagagacagaaaagagaagaaagaaagcactctgct 1438
 |||||
 Db 1499 GGTGGGTTCCAGGGACTGTACCTGTAGGAACAGAAAGAGAAAGAACACTCTGCT 1558
 |||||
 QY 1439 ggcggaatactcttgggtcaactcaaatgaagtcgggaaattctgtcttgaacttc 1498
 |||||
 Db 1559 GCGGGGAATACTCTTGGTCACTCAAAATTAAGTCGGGAAATCTGCTGTGAACATTC 1618
 |||||
 QY 1499 agccctgaacctttgt-caccattccttaattctcaaccacaaagattctctctttc 1557
 |||||
 Db 1619 AGCCCTGAACCTTTGTCACCATTCCTTTAAATTTCTCAACCCCAAGATATCTTTTTC 1678
 |||||
 QY 1558 ttagtctcaagtagctggcgtacacagcaggttaccttggctgtgctccctgtgtacc 1617
 |||||
 Db 1679 TTAGTTTCAGAAAGTACTGGCATCACAGCAGTTACCTTGGCGTGTCTCCCTGTGTTACC 1738
 |||||
 QY 1618 ctgcaagaagagacaaagcttcttccctgtctggtgccaagtcaagtaggagagtgca 1677
 |||||
 Db 1739 CGGCAGAGAAGAGACCAAGCTTTTCCCTGTCTGGCCAAAGTCACTAGGAGAGGATGCA 1798
 |||||
 QY 1678 cagtttctatttcttttagagacagagctgtataaacaagcctaactgtgtcaaaa 1737
 |||||
 Db 1799 CAGTTTGTATTCTTTAGAGACAGGAGTGTATAAACAAGCCTAACATTTGGTGCAAG 1858
 |||||
 QY 1738 a 1738
 |
 Db 1859 A 1859

RESULT 3
 V27962
 ID V27962 standard; cDNA; 1862 BP.
 AC V27962;
 DT 25-SEP-1998 (first entry)
 DE Aspartic proteinase gene ASP1.
 KW ss; ASP1; Aspartic proteinase; Alzheimer's disease; cancer; melanoma.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 91..1647
 FT /tag= a
 FT /product= "Aspartic protease ASP1"
 PN EP-848062-A2.
 PD 17-JUN-1998.
 PF 01-DEC-1997; 309648.
 PR 14-DEC-1996; GB-026022.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Chapman CG, Evans JR, Powell DJ, Southan C;
 DR WPI; 98-314477/28.
 DR P-PSDB: W61362.
 PT New isolated polynucleotide encodes Aspartic protease polypeptide -
 PT used to diagnosis, treat and vaccinate against Alzheimer's disease,
 PT cancer and melanoma
 PS Claim 2; Page 6-7; 19pp; English.
 CC The human ASP1 gene encodes a protein which is structurally related to
 CC other proteins of the Aspartic proteinase family. ASP 1 polypeptides
 CC and polynucleotides can be used to diagnosis, treat and vaccinate against
 CC Alzheimer's disease, cancer and melanoma.
 SQ Sequence 1862 BP; 386 A; 550 C; 498 G; 428 T;

Query Match 16.9%; Score 296; DB 1; Length 1862;
 Best Local Similarity 59.4%; Pred. No. 4.3e-79;
 Matches 544; Conservative 0; Mismatches 360; Indels 12; Gaps 2;

QY 53 ctggtgtgtccagcacataccgggaacctcgcggaagggtgtgtatgtgccctacacccagg 112
 |||||
 Db 485 CAGAGAGGTCTGACACATACCGCTCCAAAGGGCTTTTGACGTACAGTGAAGTACACACAAG 544
 |||||
 QY 113 gcaagtgggaaggagctgggcacccagcctggaagacatcccccatggcccaagctca 172
 |||||
 Db 545 GAAGCTGAGCGGGCTTCGTTGGGAGACCTCGTCACCATCCCCAAAGGCTTCAATACTT 604
 |||||
 QY 173 ctgtgcgtgccaaactgtgtgcccactgaatcagacaagttcttctcatcaacggctcca 232
 |||||
 Db 605 CTTTCTCTGTCAACATGCCCATAATTTTGAATCAGAGAATTTCTTTTTCCTGGGATTA 664
 |||||
 QY 233 actgggaagacacctgctgggctggcctatgtgagattgccaggcctgacgactccccgg 292
 |||||
 Db 565 AATGGAATGGAATACTTGGCCTAGCTTAGTCCACACTTGCCAAAGCCATCAAGTCTCTGG 724
 |||||
 QY 293 agcctttcttggactctctggtaaagcagacccacagcttcccaacctcttctccctcagc 352
 |||||
 Db 725 AGACCTTCTTCGACTCCCTGGTGACACAAAGCAACATCCCAACGGTTTCTCCATGCAGA 784
 |||||
 QY 353 ttgtggtgctggctcccccctcaacagctcctgaaagtgtggtcctctgtcggaggagca 412
 |||||
 Db 785 TGTGTGGAGCGGCTTGGCCGT-----TGCTGGATCTGGGACCAACGGAGGTAGTC 835
 |||||
 QY 413 tgatcattggaggtatcgaccactcgtgtacacagcagctctctgtgtatatacacccatcc 472
 |||||
 Db 836 TTCTCTTGGGTGGAATTGAACCAAGTTTGTATAAAGGAGACATCTTGGTATACCCCTATTA 895
 |||||
 QY 473 ggcggagtggttattatgaggtgactcattgtgcgggtggagatcaatggacagatctga 532
 |||||
 Db 896 AGGAAGAGTGGTACTACCAGATAGAAATTTCTGAATTTGGAATTTGGAGGCCAAGCCTTA 955
 |||||
 QY 533 aaatggactgcaaggagtacaaactatgacaagagcattgtggacagtggtgacacacaccc 592
 |||||
 Db 956 ATCTGGACTGCGAGAGAGTATAACGCAGACAAAGGCCATCGTGACAGTGGCACACCGTGC 1015
 |||||
 QY 593 ttcgtttgcacaaagaaagtgttgaagctgcagtcacaaatccatcaaggcagcctctcca 652
 |||||
 Db 1016 TCGCGCTGCCCCAGAAAGTGTGTATGCGGTGGTGGAGAGTGTGCCCGCGCATCTCTGA 1075
 |||||
 QY 653 cggagaagtccctgacggtttctgtgtagagagcagctggtgtgctgggcaagcaggca 712
 |||||
 Db 1076 TTCCAGAATTTCTGTATGTTTCTGGACTGGGTCCAGCTGGCGTGTGGAGAAATTCGG 1135
 |||||
 QY 713 ccaccccttgaacatttccacagtcactcactactaactaattgggtgagttaccaccc 772
 |||||
 Db 1136 AAACACCTTGGTCTTACTTCCCTTAAATCTCCATCTACCTGAGAGACGAGAACTCCAGCA 1195
 |||||
 QY 773 agtcttccgcataccataccttccgcagcaataactgcggccagtcggaagatgtgccca 832
 |||||
 Db 1196 GGTCAATCCGTATCACAAATCCTGCCTCAGCTTACATTCACGCCCATG---ATGCGGGCCG 1252
 |||||
 QY 833 cgtcccaagacgactgttacaagtgtgccaatcctcacagtcacccacgggacactgtatgg 892
 |||||
 Db 1253 GCCTGAATTAATGATGTACCGATTTCGGCATTTCCCATCCACAAATCCGCTGTGTATCG 1312
 |||||
 QY 893 gagctgttatcatgagggcttctacattgtcttctgtatcggtcccgaaacgaattggct 952
 |||||
 Db 1313 GTGCCAGGTGATGGAGGGCTTCTACGTCATCTTCGACAGAGCCCAAGAGGGGTGGGCT 1372
 |||||
 QY 953 ttgctgtcagcgttg 968
 |||||
 Db 1373 TCGCAGGAGGCCCTG 1388

RESULT 4

X39493

ID X39493 standard; DNA; 185 BP.

AC X39493;

DT 21-JUN-1999 (first entry)

DE Human secreted protein 5; EST SEQ ID No 91.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; hematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
OS Homo sapiens.
PN W0906551-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1235.
PR 01-AUG-1997; US-905133.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153781/13.
DR P-PSDB; Y11427.
PT New nucleic acids encoding human secreted - proteins obtained from
PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals
and fetal brain tissue
PS Claim 1; Page 209-210; 434pp; English.
CC X39440 to X39597 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y11374 to Y11531,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, hematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 185 BP; 44 A; 54 C; 51 G; 36 T; 0;

Query Match 9.5%; Score 165.4; DB 1; Length 185;
Best Local Similarity 99.4%; Pred. No. 2.5e-40;
Matches 166; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttcaccaatggtccctctcatctatctgcaagccactttacactgtcttctgggtgg 60
Db 18 GCCTCACCAGTGGTCCCTTCTATCTATCTGCAAGCCACTTACACTCTGTCTGGGTGG 77

Qy 61 tccagcacataccgggacctccggaagggtgtgtatgtgcctacacccagggaagtgg 120
Db 78 TCCAGCACATACCGGGACCTCGGAAGGGTGTGTATGTGCCCTACACCCAGGGCAAGTGG 137

Qy 121 gaaggaggagctggcaccgacctggtaagcatcccccattggccccc 167
Db 138 GAAGGGGAGCTGGGCACCGGACCTGGTAAAGCATCCCCCATGGCCCAAA 184

RESULT 5
T64547
ID T64547 standard; cDNA; 1506 BP.
AC T64547;
DE NADP-specific glutamate dehydrogenase alpha subunit cDNA.
KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW Chloroplast; transgenic plant; ds.
OS Chlorella sorokiniana strain UTEX 1230.
FH Key Location/Qualifiers
FT cds 4..1467
FT mat_peptide 4..1464
FT W09712983-A1.
PD 10-APR-1997.
PF 03-OCT-1996; U15921.
PR 06-OCT-1995; US-541033.

PA (UVFL) UNIV FLORIDA.
PI Miller P, Schmidt RR;
DR WPI: 97-226226/20.
DR P-PSDB; W15411.
PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
PT Chlorella sorokiniana algal protoplast(s), useful to modulate
PT nitrogen metabolism plant cells
PS Claim 2; Page 44-46; 61pp; English.
CC cDNA clones (T64547 and T64548) respectively code for the alpha
CC (W15411) and beta mature subunits (W15412) of ammonium-inducible,
CC chloroplast-localised NADP-specific glutamate dehydrogenase (NADP-
CC GDH) hexameric isoenzymes of Chlorella sorokiniana. They were
CC obt'd. by removal of the chloroplast targeting signal from full-
CC length cDNA clones (see also T64542-43) by PCR mutagenesis. The N
CC metabolism of plant cells can be modulated (pref. increasing them
CC assimilation of inorganic N into organic N) by transforming them
CC with nucleotide sequences encoding the alpha and/or beta mature
CC subunits or subunit precursors (see also T64529-30). Such plants
CC show improved properties, e.g. increased crop yield and improved
CC stress tolerance. Heterohexamers having alpha and beta subunits
CC can be expressed that have higher aminating/deaminating activity
CC ratios (i.e. higher capacity for glutamate synthesis) than
CC homohexamers.
SQ Sequence 1506 BP; 302 A; 495 C; 483 G; 226 T; 0;

Query Match 2.2%; Score 38.2; DB 1; Length 1506;
Best Local Similarity 49.7%; Pred. No. 0.18;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 71 accgggacctccggaagggtgtgtatgtgcccacaccaggggcaagtgggaagggaagc 130
Db 1055 ACCTGACTGCCAGGTGGACATCGCTTCCCTCGCCACCCAGACGAGATCGATGAGC 1114

Qy 131 tgggcaccgacctgtgaagcatcccccatggcccacagtcactgtgcgtgccacattg 190
Db 1115 AGGAGCGCGAGCTGCTGATCAAGCACGGGTGCCAGTACGTGTGGGGGCGCCACATGC 1174

Qy 191 ctgcatacactgaatcagacaagttcttcatacagcggtcccaactgggaaggcatcctgg 250
Db 1175 COTCCACCAAGGAGGCCATCCACAAGTAGTACAAGGCGCGCATCTACTCTCCCGGCA 1234

Qy 251 ggctggcctatgctg 265
Db 1235 AGCGGCCCAACGCCG 1249

RESULT 6
T64548
ID T64548 standard; cDNA; 1473 BP.
AC T64548;
DE NADP-specific glutamate dehydrogenase beta subunit cDNA.
KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW Chloroplast; transgenic plant; ds.
OS Chlorella sorokiniana strain UTEX 1230.
FH Key Location/Qualifiers
FT cds 4..1434
FT mat_peptide 4..1431
FT W09712983-A1.
PD 10-JUL-1997 (first entry)
PF 03-OCT-1996; U15921.
PR 06-OCT-1995; US-541033.
PA (UVFL) UNIV FLORIDA.
PI Miller P, Schmidt RR;
DR WPI: 97-226226/20.
DR P-PSDB; W15412.
PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
PT Chlorella sorokiniana algal protoplast(s), useful to modulate
PT nitrogen metabolism plant cells
PS Claim 2; Page 48-50; 61pp; English.

CC cDNA clones (T64547 and T64548) respectively-code for the alpha
 CC (W15411) and beta mature subunits (W15412) of ammonium-inducible,
 CC chloroplast-localised NADP-specific glutamate dehydrogenase (NADP-
 CC GDH) hexameric isoenzymes of Chlorella sorokiniana. They were
 CC obtd. by removal of the chloroplast targeting signal from full-
 CC length cDNA clones (see also T64542-43) by PCR mutagenesis. The N
 CC metabolism of plant cells can be modulated (pref. increasing the
 CC assimilation of inorganic N into organic N) by transforming them
 CC with nucleotide sequences encoding the alpha and/or beta mature
 CC subunits or subunit precursors (see also T64529-30). Such plants
 CC show improved properties, e.g. increased crop yield and improved
 CC stress tolerance. Heterohexamers having alpha and beta subunits
 CC can be expressed that have higher aminating:deaminating activity
 CC ratios (i.e. higher capacity for glutamate synthesis) than
 CC homohexamers.
 SQ Sequence 1473 BP; 297 A; 486 C; 470 G; 220 T;

Query Match 2.1%; Score 36.6; DB 1; Length 1473;
 Best Local Similarity 49.2%; Pred. No. 0.53;
 Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 71 accgggacctccggaagggtgtgtatgtgccctacaccagggcaagtgggaagggaagc 130
 DB 1022 AGCTGGACTGCGAGGTGGACATCGCTTCCCTGCGCCACCCAGACGATCGATGAGC 1081
 QY 131 tgggcaccgacctgtaagcatccccatggcccccaagtcactgtgctgccaacattg 190
 DB 1082 ACGACGCCGAGCTGTGATCAAGCACGCGTGCAGTAGTACGAGCGCGCCCAACATGC 1141
 QY 191 ctgcatcactgaatcagacaagtcttctcatcacgcgtccaaactgggaaggcatctctg 250
 DB 1142 CCTCCACCAACGAGGCCATCCACAAGTACACAAGCGCGCATCTACTGCCCCGGCA 1201
 QY 251 ggtcgccatgtctg 265
 DB 1202 AGCGGCCCAACGCCG 1216

RESULT 7
 ID T64531 standard; cDNA; 1969 BP.
 AC T64531; 1997 (first entry)
 DE Glutamate dehydrogenase cDNA clone pBGdc53.
 KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
 KW Chloroplast; transgenic plant; ds.
 OS Chlorella sorokiniana strain UTEX 1230.
 PN W09712983-Al.
 PD 10-APR-1997.
 PF 03-OCT-1996; U15921.
 PR 06-OCT-1995; US-541033.
 PA (UYFL) UNIV FLORIDA.
 PI Miller P, Schmidt RR;
 DR WPI: 97-226226/20.
 PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
 PT Chlorella sorokiniana algal protoplast(s), useful to modulate
 PT nitrogen metabolism plant cells
 PS Claim 2; Page 35-36; 61pp; English.
 CC A cDNA clone (T64531), designated pBGdc53, was isolated from a
 CC Chlorella sorokiniana cDNA library in lambda ZAP by screening
 CC with an NADP-specific glutamate dehydrogenase (NADP-GDH) cDNA
 CC probe. It comprises the complete 3'-terminal region of the
 CC C. sorokiniana NADP-GDH sequence. 5' terminal regions were
 CC obtd. by PCR, yielding full-length clones (T64542-43) for
 CC the alpha and beta subunits of C. sorokiniana NADP-GDH.
 SQ Sequence 1969 BP; 401 A; 659 C; 580 G; 329 T;

Query Match 2.1%; Score 36.6; DB 1; Length 1969;
 Best Local Similarity 49.2%; Pred. No. 0.62;
 Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 71 accgggacctccggaagggtgtgtatgtgccctacaccagggcaagtgggaagggaagc 130
 DB 1031 AGCTGGACTGCGAGGTGGACATCGCTTCCCTGCGCCACCCAGACGATCGATGAGC 1090
 QY 131 tgggcaccgacctgtaagcatccccatggcccccaagtcactgtgctgccaacattg 190
 DB 1091 ACGACGCCGAGCTGTGATCAAGCACGCGTGCAGTAGTACGAGCGCGCCCAACATGC 1150
 QY 191 ctgcatcactgaatcagacaagtcttctcatcacgcgtccaaactgggaaggcatctctg 250
 DB 1151 CCTCCACCAACGAGGCCATCCACAAGTACACAAGCGCGCATCTACTGCCCCGGCA 1210
 QY 251 ggtcgccatgtctg 265
 DB 1211 AGCGGCCCAACGCCG 1225

RESULT 8
 ID T64543 standard; cDNA; 2096 BP.
 AC T64543;
 DE NADP-specific glutamate dehydrogenase beta-subunit precursor cDNA.
 KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
 KW Chloroplast; transgenic plant; ds.
 OS Chlorella sorokiniana strain UTEX 1230.
 FH Key Location/Qualifiers
 FT cds 30..1568
 FT transit_peptide 30..146 /*tag= a
 FT mat_peptide 147..1565 /*tag= b
 FT /*tag= c
 PN W09712983-Al.
 PD 10-APR-1997.
 PF 03-OCT-1996; U15921.
 PR 06-OCT-1995; US-541033.
 PA (UYFL) UNIV FLORIDA.
 PI Miller P, Schmidt RR;
 DR WPI: 97-226226/20.
 DR P-PSDB; W15408.
 PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
 PT Chlorella sorokiniana algal protoplast(s), useful to modulate
 PT nitrogen metabolism plant cells
 PS Claim 2; Page 41-43; 61pp; English.
 CC cDNA clones (T64542 and T64543) respectively code for the alpha
 CC subunit (W15407) and beta subunit (W15408) precursor proteins of an
 CC ammonium-inducible, chloroplast-localised NADP-specific glutamate
 CC dehydrogenase (NADP-GDH) of Chlorella sorokiniana. They were obtd.
 CC by ligation of 3' clone pBGdc53 (T64531) and 5' clones pBGdc63
 CC (T64540) and pBGdc64 (T64541). The proteins are processed to yield
 CC mature alpha and beta subunits (see also T64547-48) that comprise
 CC the active NADP-GDH hexameric isoenzymes. The nitrogen metabolism
 CC of plants can be modulated (pref. by increasing assimilation of
 CC inorganic N into organic N) by transforming them with nucleotide
 CC sequences encoding the alpha and/or beta subunits. Such plants show
 CC improved properties, e.g. increased crop yield and improved stress
 CC tolerance. Heterohexamers having alpha and beta subunits can be
 CC expressed that have higher aminating:deaminating activity ratios
 CC (i.e. higher capacity for glutamate synthesis) than homohexamers.
 SQ Sequence 2096 BP; 416 A; 711 C; 616 G; 353 T;

Query Match 2.1%; Score 36.6; DB 1; Length 2096;
 Best Local Similarity 49.2%; Pred. No. 0.65;
 Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 71 accgggacctccggaagggtgtgtatgtgccctacaccagggcaagtgggaagggaagc 130
 DB 1156 AGCTGGACTGCGAGGTGGACATCGCTTCCCTGCGCCACCCAGACGATCGATGAGC 1215
 QY 131 tgggcaccgacctgtaagcatccccatggcccccaagtcactgtgctgccaacattg 190

Db 1216 ACGACGCCGAGCTGCTGATCAAGCAGCGCTGCCAGTACGTGGTGGAGGCGGCCAACATGC 1275
QY 191 ctgccatcaactgaatcagacaagttcttcatacagcggtcccaactgggaagcattccctgg 250
Db 1276 CTTCCACCAACAGGCCATCCACAAGTACACAGAGCGCGGCATCATCTACTGCCCGCGCA 1335
QY 251 ggctggcctatgctg 265
Db 1336 AGCGGCCCAACGCCG 1350

RESULT 9

T64530
ID T64530 standard; cDNA; 2099 BP.
AC T64530;
DE 10-JUL-1997 (first entry)
DE NADP-specific glutamate dehydrogenase beta-subunit precursor cDNA.
KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW Chloroplast; transgenic plant; ds.
OS Chlorella sorokiniana strain UTEX 1230.
FH Key Location/Qualifiers
FT cds 33..1571
FT
FT transit_peptide 33..140 /*tag= a
FT
FT mat_peptide 141..1568 /*tag= b
FT
FT
FT
PN WO9712983-A1.
PD 10-APR-1997.
PF 03-OCT-1996; U15921.
PR 06-OCT-1995; US-541033.
PA (UYFL) UNIV FLORIDA.
PI Miller P, Schmidt RR;
DR WPI; 97-226226/20.
DR P-PSDB; W15408.

PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
PT Chlorella sorokiniana algal protoplast(s), useful to modulate
PT nitrogen metabolism plant cells
PS Claim 2: Page 29-32; 61pp; English.
CC cDNA clones (T64529 and T64530) respectively code for the alpha
CC subunit (W15407) and beta subunit (W15408) precursor proteins of an
CC ammonium-inducible, chloroplast-localised NADP-specific glutamate
CC dehydrogenase (NADP-GDH) of Chlorella sorokiniana. They were
CC deposited in E. coli DH5alpha as ATCC 69925 and 69926, respectively.
CC The precursor proteins are processed to mature alpha and beta
CC subunits (see also T64547-48) that comprise active NADP-GDH
CC hexameric isoenzymes. The nitrogen metabolism of plants can be
CC modulated (pref. increasing the assimilation of inorganic nitrogen
CC into organic nitrogen) by transforming them with nucleotide sequences
CC encoding the alpha and/or beta subunits. Such plants show improved
CC properties, e.g. increased crop yield and improved stress tolerance.
CC Heterohexamers having alpha and beta subunits can be expressed that
CC have higher aminating/deaminating activity ratios (i.e. higher
CC capacity for glutamate synthesis) than homohexamers.
SQ Sequence 2099 BP; 416 A; 713 C; 616 G; 354 T;

Query Match 2.1%; Score 36.6; DB 1; Length 2099;
Best Local Similarity 49.2%; Pred. No. 0.65;
Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 71 accgggacctccgaaggggtgtatgtgcccctacaccagggccaagtgggaagggagc 130
Db 1159 AGCTGGATGCGCAGGTGGACATCGCTTCCCTGGCCACCAGAACGAGATCGATGAGC 1218
QY 131 tgggcaccgacctggaagcattccccatggcccccaacgtcactgtgcgtgccaacattg 190
Db 1219 ACGAGCGCGAGCTGCTGATCAAGCAGCGCTGCCAGTACGTGGTGGAGGCGGCCACATGC 1278
QY 191 ctgccatcaactgaatcagacaagttcttcatacagcggtcccaactgggaagcattccctgg 250
Db 1279 CTTCCACCAACAGGCCATCCACAAGTACACAGAGCGCGGCATCATCTACTGCCCGCGCA 1338

QY 251 ggctggcctatgctg 265
Db 1339 AGCGGCCCAACGCCG 1353

RESULT 10

T64542
ID T64542 standard; cDNA; 2137 BP.
AC T64542;
DE 10-JUL-1997 (first entry)
DE NADP-specific glutamate dehydrogenase alpha-subunit precursor cDNA.
KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW Chloroplast; transgenic plant; ds.
OS Chlorella sorokiniana strain UTEX 1230.
FH Key Location/Qualifiers
FT cds 30..1610
FT
FT transit_peptide 30..149 /*tag= a
FT
FT mat_peptide 150..1607 /*tag= b
FT
FT
FT
PN WO9712983-A1.
PD 10-APR-1997.
PF 03-OCT-1996; U15921.
PR 06-OCT-1995; US-541033.
PA (UYFL) UNIV FLORIDA.
PI Miller P, Schmidt RR;
DR WPI; 97-226226/20.
DR P-PSDB; W15407.
PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
PT Chlorella sorokiniana algal protoplast(s), useful to modulate
PT nitrogen metabolism plant cells
PS Claim 2: Page 40-41; 61pp; English.
CC cDNA clones (T64542 and T64543) respectively code for the alpha
CC subunit (W15407) and beta subunit (W15408) precursor proteins of an
CC ammonium-inducible, chloroplast-localised NADP-specific glutamate
CC dehydrogenase (NADP-GDH) of Chlorella sorokiniana. They were obtd.
CC by ligation of 3' clone pBGdc53 (T64531) and 5' clones pBGdc63
CC (T64540) and pBGdc64 (T64541). The proteins are processed to yield
CC mature alpha and beta subunits (see also T64547-48) that comprise
CC the active NADP-GDH hexameric isoenzymes. The nitrogen metabolism
CC of plants can be modulated (pref. by increasing assimilation of
CC inorganic N into organic N) by transforming them with nucleotide
CC sequences encoding the alpha and/or beta subunits. Such plants show
CC improved properties, e.g. increased crop yield and improved stress
CC tolerance. Heterohexamers having alpha and beta subunits can be
CC expressed that have higher aminating/deaminating activity ratios
CC (i.e. higher capacity for glutamate synthesis) than homohexamers.
SQ Sequence 2137 BP; 416 A; 732 C; 632 G; 357 T;

Query Match 2.1%; Score 36.6; DB 1; Length 2137;
Best Local Similarity 49.2%; Pred. No. 0.65;
Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 71 accgggacctccgaaggggtgtatgtgcccctacaccagggccaagtgggaagggagc 130
Db 1198 AGCTGGATGCGCAGGTGGACATCGCTTCCCTGGCCACCAGAACGAGATCGATGAGC 1257
QY 131 tgggcaccgacctggaagcattccccatggcccccaacgtcactgtgcgtgccaacattg 190
Db 1258 ACGAGCGCGAGCTGCTGATCAAGCAGCGGTGCCAGTACGTGGTGGAGGCGGCCACATGC 1317
QY 191 ctgccatcaactgaatcagacaagttcttcatacagcggtcccaactgggaagcattccctgg 250
Db 1318 CTTCCACCAACAGGCCATCCACAAGTACACAAGCGCGGCATCATCTACTGCCCGCGCA 1377
QY 251 ggctggcctatgctg 265
Db 1378 AGCGGCCCAACGCCG 1392

RESULT 11

T64529
 ID T64529 standard; cDNA; 2140 BP.
 AC T64529;
 DT 10-JUL-1997 (first entry)
 DE NADP-specific glutamate dehydrogenase alpha-subunit precursor cDNA.
 KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
 KW chloroplast; transgenic strain; ds.
 OS Chlorella sorokiniana strain UTEX 1230.
 FT Key Location/Qualifiers
 FT cds 33..1613
 FT transit_peptide 33..152 /*tag= a
 FT mat_peptide 153..1610 /*tag= b
 FT /*tag= c
 FN WO9712983-Al.
 PD 10-APR-1997.
 PF 03-OCT-1996; U15921.
 PR 06-OCT-1995; US-541033.
 PA (UYFL) UNIV FLORIDA.
 PI Miller P, Schmidt RR;
 DR WPI: 97-226226/20.
 DR P-PSDB; W15407.
 PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
 PT Chlorella sorokiniana algal protoplast(s), useful to modulate
 PT nitrogen metabolism plant cells
 PS Claim 2; Page 25-27; 61pp; English.
 CC cDNA clones (T64529 and T64530) respectively code for the alpha
 CC subunit (W15407) and beta subunit (W15408) precursor proteins of an
 CC ammonium-inducible, chloroplast-localised NADP-specific glutamate
 CC dehydrogenase (NADP-GDH) of Chlorella sorokiniana. They have been
 CC deposited in E. coli DH5alpha as ATCC 69925 and ATCC 69926,
 CC respectively. The precursor proteins are processed to yield mature
 CC alpha and beta subunits (see also T64547-48) that comprise the
 CC active NADP-GDH hexameric isoenzymes. The nitrogen metabolism of
 CC plants can be modulated (pref. increasing the assimilation of
 CC inorganic N into organic N) by transforming them with nucleotide
 CC sequences encoding the alpha and/or beta subunits. Such plants show
 CC improved properties, e.g. increased crop yield and improved stress
 CC tolerance. Heterohexamers having alpha and beta subunits can be
 CC expressed that have higher aminating/deaminating activity ratios
 CC (i.e. higher capacity for glutamate synthesis) than homohexamers.
 SQ Sequence 2140 BP; 416 A; 734 C; 632 G; 358 T;

Query Match 2.1%; Score 36.6; DB 1; Length 2140;
 Best Local Similarity 49.2%; Pred. No. 0.65;
 Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
 QY 71 acggggacctccggaaggtgtgtatgtgcccctacacccaggggcaagtgggaagggagc 130
 Db 1201 AGCTGGACTCCAGGTGGACATGCGCTTCCCTGGCCGCCACCCAGACGATCGATGAGC 1260
 QY 131 tgggacccagcctggtaagcatcccccatgcccccaactgacctgtgcgaacattg 190
 Db 1261 ACGACGCCGAGCTGCTGATCAACACGCGGTGCGCATGCTGGAGGGCGCCACATGC 1320
 QY 191 ctgccatcactgacagacaagtcttcatcaacggctcccaactgggaaggaagcattcg 250
 Db 1321 CCTCCACCAACGAGGCCATCCACAAAGTACAAACAGCCCGGCATCATCTACTGCCCGCGCA 1380
 QY 251 ggctggcctatctg 265
 Db 1381 AGCGGCCCAACGCCG 1395

RESULT 12

V59498
 ID V59498 standard; cDNA; 2288 BP.
 AC V59498;
 DT 02-FEB-1999 (first entry)
 DE Human sodium dependent phosphate transporter IPT-1 cDNA.
 KW IPT-1; sodium dependent phosphate transporter; human;

KW kidney failure; kidney disease; uraemic bone disease; cancer;
 KW diagnosis; gene therapy; vaccine; ss.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT cds 64..2136
 FT /*tag= a
 FN EP-875569-Al.
 PD 04-NOV-1998.
 PF 09-APR-1998; 302815.
 PR 23-SEP-1997; US-935433.
 PR 28-APR-1997; US-044974.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Feild J;
 DR WPI: 98-559435/48.
 DR P-PSDB; W77414.
 PT New DNA encoding sodium-dependent phosphate transporter protein
 PT IPT-1 - used to prevent, treat and diagnose e.g. chronic renal
 PT failure, end-stage renal disease, uraemic bone disease and cancer
 PS Claim 2; Page 16-17; 24pp; English.
 CC This isolated polynucleotide includes an open reading frame coding
 CC for human sodium dependent phosphate transporter IPT-1 (see
 CC W77414). Polynucleotides encoding IPT-1 can be obtained e.g. from
 CC a cDNA library derived from mRNA in cells of human small intestine
 CC and lung using expressed sequence tag analysis, or can be obtained
 CC from genomic DNA libraries. The nucleotide sequence provided has
 CC about 78% identity in 2242 nucleotide residues with bovine sodium
 CC dependent phosphate transporter and 58% identity in 2270 nucleotides
 CC with human Napi-3. IPT-1 polynucleotides of the invention can be
 CC used in the recombinant production of IPT-1 polypeptides. They are
 CC also used in claimed methods for treating a subject in need of
 CC enhanced activity or expression of IPT-1, or (e.g. antisense) for
 CC treating a subject having need to inhibit IPT-1 activity or
 CC expression. A claimed process for diagnosing a disease or
 CC susceptibility to disease related to IPT-1 expression or activity
 CC involves determining the presence or absence of a mutation in the
 CC IPT-1 nucleotide sequence of a subject. IPT-1 polypeptides and
 CC polynucleotides can be used for prevention, treatment and diagnosis
 CC of e.g. chronic renal failure, end-stage renal disease, uraemic
 CC bone disease and cancer.
 SQ Sequence 2288 BP; 488 A; 672 C; 575 G; 553 T;

Query Match 2.0%; Score 35.6; DB 1; Length 2288;
 Best Local Similarity 60.2%; Pred. No. 1.4;
 Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1090 tgtcatggtccatctgcgcccctcttcctatgtgccactctgctcatgtgtgtcagtg 1149
 Db 1734 TGTGGGGGTTTCCCGTCGTTCATCATCATCTGTGCTCGACTCTCTGCACTC 1793
 QY 1150 gcgctgctccctgcctgcgcagcagcatgatgact 1187
 Db 1794 TCGCTGCCAGCGCTCTCTCGGAAGAACTCCAGAACT 1831

RESULT 13

V62392/c
 ID V62392 standard; DNA; 12565 BP.
 AC V62392;
 DT 19-JAN-1999 (first entry)
 DE Human interleukin-1 receptor antagonist gene.
 KW Human; interleukin-1 receptor antagonist gene; IL-1; polymorphism;
 KW diagnosis; osteoporosis; ds.
 OS Homo sapiens.
 FN WO9844150-Al.
 PD 08-OCT-1998.
 PF 27-MAR-1998; G00944.
 PR 27-MAR-1997; GB-006359.
 PA (GEMT-) GEMINI RES LTD.
 PI Keen RW, Spector TD;
 DR WPI: 98-557135/47.
 PT Diagnosis of osteoporosis by determining genotype of interleukin-1
 PT receptor antagonist gene - useful for diagnosing patient

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 21:01:28 ; Search time 203.6 seconds
(without alignments)
1115.343 Million cell updates/sec

Title: US-09-215-435-66

Perfect score: 1747

Sequence: 1 gccctcacaatggtccctt.....tggtgcacaaaaaaaaa 1747

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2.6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2.6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2.6/ptodata/1/ina/5C_COMB.seq:*
- 4: /cgn2.6/ptodata/1/ina/5D_COMB.seq:*
- 5: /cgn2.6/ptodata/1/ina/6_COMB.seq:*
- 6: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq:*
- 7: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296	16.9	1862	5	US-08-999-723-1
2	36.6	2.1	1473	3	US-08-541-033A-25
3	36.6	2.1	1473	4	US-08-828-451-25
4	36.6	2.1	1506	3	US-08-541-033A-23
5	36.6	2.1	1506	4	US-08-828-451-23
6	36.6	2.1	1369	3	US-08-541-033A-7
7	36.6	2.1	1369	4	US-08-828-451-7
8	36.6	2.1	2096	3	US-08-541-033A-19
9	36.6	2.1	2096	4	US-08-828-451-19
10	36.6	2.1	2099	3	US-08-541-033A-3
11	36.6	2.1	2099	4	US-08-828-451-3
12	36.6	2.1	2137	3	US-08-541-033A-18
13	36.6	2.1	2137	4	US-08-828-451-18
14	36.6	2.1	2140	3	US-08-541-033A-1
15	36.6	2.1	2140	4	US-08-828-451-1
16	35.4	2.0	276	1	US-08-127-954-98
17	35.4	2.0	276	1	US-08-127-954-108
18	35.4	2.0	276	1	US-08-127-954-109
19	35.4	2.0	276	1	US-08-127-954-110
20	35.4	2.0	276	1	US-08-127-954-111
21	33.8	1.9	276	1	US-08-127-954-95
22	33.8	1.9	276	1	US-08-127-954-96
23	33.8	1.9	276	1	US-08-127-954-97
24	33.8	1.9	276	1	US-08-127-954-99
25	33.8	1.9	276	1	US-08-127-954-100
26	33.8	1.9	276	1	US-08-127-954-102
27	33.8	1.9	276	1	US-08-127-954-103

ALIGNMENTS

RESULT 1

US-08-999-723-1
; Sequence 1, Application US/08999723A
; Patent No. 6025180

; GENERAL INFORMATION:
; APPLICANT: Powell, David J.

; APPLICANT: Southan, Christopher

; APPLICANT: Chapman, Conrad G.

; APPLICANT: Evans, Joanne R.

; TITLE OF INVENTION: ASP1

; FILE REFERENCE: GH70262

; CURRENT APPLICATION NUMBER: US/08/999,723A

; CURRENT FILING DATE: 1997-10-06

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1862

; TYPE: DNA

; ORGANISM: Homo sapiens

US-08-999-723-1

Query Match 16.9%; Score 296; DB 5; Length 1862;
Best Local Similarity 59.4%; Pred. No. 9.8e-83;
Matches 544; Conservative 0; Mismatches 360; Indels 12; Gaps 2;

QY	53	ctgggtggtccagcacatccgggaccccggaagggtgtgtatgtgccctacacccagg	112
Db	485	cadagaggttagcacatcccgctcccaagggtttgacgtcacagtgaagtacacaaag	544
QY	113	gcaagtgggaaggagctgggcacccgacctgtaagcatcccccattggcccacagtc	172
Db	545	gaagctggagcggccttctgtggggaagacctgtcacatcccccaggcttcaactt	604
QY	173	ctgtggtgccaacattgctgccatcacctgaatcagacagtttctcatcaacgctcca	232
Db	605	ctttctgtcaacattgccactattttgaatcaggaatttttttggctgggatta	664
QY	233	actgggaagcattctgggctgcttatgctgagattgcccagcctgagactccccgg	292
Db	665	aatgggaatgaatacttggctgcttatgccacacttgcacagccatcaagtctctgg	724
QY	293	agcctttctgactctctgttaagcagaccacgcttcccaacctcttctccctgagc	352
Db	725	agaccttctgactctctgttgaacaaacacacacacacacacacacacacacacac	784
QY	353	tttgtgtgctgcttccctccctcaacagctctgaagtgtgcttctgtcgaggagga	412
Db	785	tgtgtgagcggcgttgcctgt-----tgcgtatctgggacacacgaggtagtc	835

QY 413 tgatcattggagggtatcgaccactcgtgtacacaggcagctctctgtgtatcacaccatcc 472
Db 836 ttgtctgggtggaattgaaccaagttgtataaaggagacatctgtgtataccctatta 895
QY 473 ggcggagtgattatgagtgatctgtgcgggtggagatcaattggacagagatctga 532
Db 896 aggaagagtgctaccagataaattctgaaattggaaattggaggccaaagcctta 955
QY 533 aaatggactgcaaggtacacattatgacaagagcattgtggacagtgccaccacaacc 592
Db 956 atctggactgcagagataaagcagacagagccatcgtggacagtgccaccacgtgc 1015
QY 593 ttctgttgcacaaagagtggttgaagctgcagtcacaaatccatcaaggcagctctccca 652
Db 1016 tgcgctgcccagaagtggttgcggtggtggaagctgtgcccgcgcacatctga 1075
QY 653 cggagaggttccctgcagcttctgcgttagagagcagctggtgtgctgcaagcaggca 712
Db 1076 ttccagaaattctctgagtggttctgcgactgggtcccgagctggcgtgtgagcaattcgg 1135
QY 713 ccacccttggaacattttcccgatcactcactctacctaattgggtgaggttaccaacc 772
Db 1136 aacaccttggttacttccctaaatctccatctacctgcagagacagaaactccagca 1195
QY 773 agtcttccgcataccatcttccgcagcaaatcctgcgagtgccagtggaagatgtggcca 832
Db 1196 ggtcattccgtatcaacaatctgcctcagcttccattacattcagccatg---atggggggcg 1252
QY 833 cgtcccaagcagactgttacaagtttgccatctcagctacacagctacacagggcactgttatgg 892
Db 1253 gctgaattatgaattaccgattcggcatttcccatcccaaatgcgctggtgatcg 1312
QY 893 gactgttatcaggggcttctacgttctgttctgttgatggggccgaaacgaattggct 952
Db 1313 gtgccacggtgaggggtctctacgtctcttcacgtctcttcacagagccagagaggggtggtct 1372
QY 953 ttgctgtcagcgcttg 968
Db 1373 tgcagcgagccctg 1388

RESULT 2

US-08-541-033A-25
; Sequence 25, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541.033A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..1431
; US-08-541-033A-25

Query Match 2.1%; Score 36.6; DB 3; Length 1473;
Best Local Similarity 49.2%; Pred. No. 0.15;

Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 71 acggggacctcggaaagggtgtgtatgtgcctacacccagggcaagtgggaaggagc 130
Db 1022 AGCTGACTGCCAGGTGGACATCGCCTTCCCTGCGCCACCCAGAACGAGATCGATGAGC 1081
QY 131 tgggcaccgacctggttaagcatcccccatgcccacagctcactgtgctgccaacattg 190
Db 1082 ACGACGCCGAGCTGCTGATCAAGACAGCGGCTGCCAGTACGTGGTGGAGGCGCCAAATGC 1141
QY 191 ctgcatactactgaatcagacaagttcttctcaacaggtctccaaactgggaaggcactcctgg 250
Db 1142 CTTCCACCAACAGGCGCATCCACAAGTACAAACAGCGCGGCATCATCTACTGCCCCGGCA 1201
QY 251 ggctggcctatgctg 265
Db 1202 AGCGCGCAACGCCG 1216

RESULT 3

US-08-828-451-25
; Sequence 25, Application US/08828451
; Patent No. 5985634
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 1473 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 4...1431

US-08-828-451-25

Query Match

Best Local Similarity 2.1%; Score 36.6; DB 4; Length 1473;

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 71 accgggacccctccggaagggtgtgtatgtgcccctacacccaggggcaagtgggaaggggagc 130

Db 1022 AGCTGGACTGCCAGGTGCATCGCCCTCCCTGCGCCACCCAGACGAGATCGATGAGC 1081

QY 131 tgggacccgacctggtaagcatcccccatgcccccaacgctcaactgtgctgccaacattg 190

Db 1082 ACGACGCCGAGCTGCTGATCAAGCACGCGTCCCGTGGCGCCACCCAGACGAGATCGATGAGC 1141

QY 191 ctgccatcactgaatcagacaagttcttcacacggtcccaactggaaggcatcctgg 250

Db 1142 CCTCCACACGAGGCCATCCACNAAGTACACACAGGCGCGGCATCATCTACTGCCCGGCA 1201

QY 251 ggctggcctatgctg 265

Db 1202 AGCGGCCAACGCCG 1216

RESULT 4

US-08-541-033A-23

Sequence 23, Application US/08541033A

Patent No. 5879941

GENERAL INFORMATION:

APPLICANT: Schmidt, Robert R.

APPLICANT: Miller, Philip

TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES

TITLE OF INVENTION: RELATING TO THE '-' AND -SUBUNITS OF GLUTAMATE

TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: Florida

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/541,033A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.

REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UF155

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 1506 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 4...1464

US-08-541-033A-23

Query Match

Best Local Similarity 2.1%; Score 36.6; DB 3; Length 1506;

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 71 accgggacccctccggaagggtgtgtatgtgcccctacacccaggggcaagtgggaaggggagc 130

Db 1055 AGCTGGACTGCCAGGTGCATCGCCCTCCCTGCGCCACCCAGACGAGATCGATGAGC 1114

QY 131 tgggacccgacctggtaagcatcccccatgcccccaacgctcaactgtgctgccaacattg 190

Db 1115 ACGACGCCGAGCTGCTGATCAAGCACGCGTCCCGTGGCGCCACCCAGACGAGATCGATGAGC 1174

QY 191 ctgccatcactgaatcagacaagttcttcacacggtcccaactggaaggcatcctgg 250

Db 1175 CCTCCACACGAGGCCATCCACAAAGTACAAAGCGCGGCATCATCTACTGCCCGGCA 1234

QY 251 ggctggcctatgctg 265

Db 1235 AGCGGCCAACGCCG 1249

RESULT 5

US-08-828-451-23

Sequence 23, Application US/08828451

Patent No. 5985634

GENERAL INFORMATION:

APPLICANT: Schmidt, Robert R.

APPLICANT: Miller, Philip

TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES

TITLE OF INVENTION: RELATING TO THE '-' AND -SUBUNITS OF GLUTAMATE

TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: Florida

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/828,451

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/541,033

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.

REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UF155

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 1506 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..1464
; US-08-828-451-23

Query Match          2.1%; Score 36.6; DB 4; Length 1506;
Best Local Similarity 49.2%; Pred. No. 0.15;
Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 71 accgggacctccggaagggtgtgtatgtgcccctacacccaggggcaagtgggaaggaggc 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1055 AGCTGACTGCCAGTGCAGATGCCCTTCCCTGCGGCCACCCAGAGAGATCGATGAGC 1114

Qy 131 tgggcaccgacctggaagcattcccccatggcccccaactgctgctgcccacattg 190
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1115 ACGAGCGCGAGCTGCTGATCAAGCACGGCTGCCAGTAGCTGGTGGAGGGCGCCACATGC 1174

Qy 191 ctgcatcactgaatcagacagttcttcatacagcgtcctcaactggaaggatcctgg 250
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1175 CTTCCACCAAGAGGGCCATCCACAAGTACAAACAGGCGCGCATCTACTGCCCCGGCA 1234

Qy 251 ggcctggcctatgctg 265
    | | | | | | | |
Db 1235 AGCGGCCAACGCCG 1249

RESULT 6
US-08-541-033A-7
; Sequence 7, Application US/08541033A
; Patent No. 587941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE \- AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UFI155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-541-033A-7

Query Match          2.1%; Score 36.6; DB 3; Length 1969;
Best Local Similarity 49.2%; Pred. No. 0.18;
Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 71 accgggacctccggaagggtgtgtatgtgcccctacacccaggggcaagtgggaaggaggc 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1031 AGCTGGACTGCCAGTGCAGATGCCCTTCCCTGCGGCCACCCAGAGAGATCGATGAGC 1090

Qy 131 tgggcaccgacctggaagcattcccccatggcccccaactgctgctgcccacattg 190
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1091 ACGAGCGCGAGCTGCTGATCAAGCACGGCTGCCAGTAGCTGGTGGAGGGCGCCACATGC 1150

Qy 191 ctgcatcactgaatcagacagttcttcatacagcgtcctcaactggaaggatcctgg 250
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1151 CTTCCACCAAGAGGGCCATCCACAAGTACAAACAGGCGCGCATCTACTGCCCCGGCA 1210

Qy 251 ggcctggcctatgctg 265
    | | | | | | | |
Db 1211 AGCGGCCAACGCCG 1225

RESULT 7
US-08-828-451-7
; Sequence 7, Application US/08828451
; Patent No. 5985634
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE \- AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UFI155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-828-451-7

Query Match          2.1%; Score 36.6; DB 4; Length 1969;
Best Local Similarity 49.2%; Pred. No. 0.18;
Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 71 accgggacctccggaagggtgtgtatgtgcccctacacccaggggcaagtgggaaggaggc 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1031 AGCTGGACTGCCAGTGCAGATGCCCTTCCCTGCGGCCACCCAGAGAGATCGATGAGC 1090

```

RESULT 9
US-08-828-451-19
; Sequence 19, Application US/08828451

```

; REFERENCE/DOCUMENT NUMBER: 06-193
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (904) 375-8100
;
; TELEFAX: (904) 372-5800
;
; INFORMATION FOR SEQ ID NO: 19:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 2096 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;     MOLECULE TYPE: CDNA
; US-08-828-451-19

```

	Db	1156	AGCTGGACTGCCAGGTGGACATCGCCTTCCCTGCGCCACCAGAACGAGATCGATGAGC	12115
	Qy	131	tgggaccagcacttgtaacgcataccccctatggcccaacgtcaactgtgcgtgccaaacattg	190
/	Qy	1216	ACGACGCCGAGCTGCCTGATCAAGCAGCGGTGCCAGTAGCTGGTGAGGGCGCCAACATGC	1275
	Qy	191	ctgccataactgaatcacagacaagtcttcatcaaaggctccaacttgggaagggcataccttgg	250
	Db	1276	CCTCCACAACGAGGGCATCCACAAGTAGTAACAAGSCGGCATCATCTACTGCCCCGCCA	1335
	Qy	251	ggctggcctatgctg	265
	Db	1336	AGGCGGCCAACGCCG	1350

```

RESULT 10
US-08-541-033A-3
; Sequence 3, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE '- AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DEHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..1568
; IS-08-541-033A-3

```

[illegible]

RESULT 11
US-08-828-451-3
; Sequence 3, Application US/08828451
; Patent No. 5985634

```

; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
; APPLICANT: Miller, Philip
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; TITLE OF INVENTION: RELATING TO THE - AND -SUBUNITS OF GLUTAMATE
; TITLE OF INVENTION: DRYHYDROGENASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ. ID NO.: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..1568
; US-08-828-451-3

Query Match          2.1%; Score 36.6; DB 4; Length 2099;
Best Local Similarity 49.2%; Pred.No. 0.19;
Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps

Qy 71 accggagacctcgggaagggtgtatgtgccctacaccaggggcaagtgggaaggggc 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1159 AGTGGAGTGCCAGGTGGACATCGCTTCCTCGGCCACCAGACGATCGATGAGC 1218
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 131 tgggcccacacttgtaaacatcccccatcgccccaagtcactgtcggtcccaactg 190
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1219 ACGACGCCGAGCTGCTGATCAAGACGCGCTGCAGTAGCTGTGGAGGGCCCAACATGC 1278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 191 ctgccatcactgaatacagaagaagtcttcatcaacggctccaactggaaagcatctgg 250
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1279 CCTCCACCAACGAGGCCATCCACAAGTACACAGGCGCGCATCTACTGCCCCCGGA 1338
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 251 ggctggcctatgctg 265
   | | | | | | | | | |
Db 1339 AGGCGGCCAACGCCG 1353
   | | | | | | | | | |

RESULT 12
US-08-541-033A-18
; Sequence 18, Application US/08541033A
; Patent No. 5879941
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Robert R.
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 21:03:48 ; Search time 3550.87 Seconds
(without alignments)
-2051.155 Million cell updates/sec

Title: US-09-215-435-66
Perfect score: 1747
Sequence: 1 gctcaccatggttcctt.....tggtgcaaaaaaaaaaaaaa 1747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues
Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA:*
1: /cgn2_6/ptodata/1/pna/PCNUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq.*

45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US095_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US60001A_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US60001B_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US60001C_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US60002A_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US60002B_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US60002C_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US60003A_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US60003B_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US60004A_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US60004B_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US60004C_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US60005_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US60006_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US60007_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US60008A_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US60008B_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US60008C_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US60009A_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US60009B_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq.*
86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq.*
87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq.*
88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq.*
90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq.*
91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq.*
92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq.*
93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq.*
94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq.*
95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq.*
96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq.*
97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq.*
98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	IDB	Description
1	1746	99.9	1747	43	US-09-215-435-66 Sequence 66, Appl
2	1699.6	97.3	1773	76	US-60-096-116-57 Sequence 57, Appl
3	1644.8	94.1	2070	57	US-09-548-365-3 Sequence 3, Appl

4 1644.8 94.1 2070 57 US-09-548-366-3 Sequence 3, Appl1
5 1644.8 94.1 2070 57 US-09-548-367-3 Sequence 3, Appl1
6 1644.8 94.1 2070 57 US-09-548-368-3 Sequence 3, Appl1
7 1644.8 94.1 2070 57 US-09-548-372-3 Sequence 3, Appl1
8 1644.8 94.1 2070 57 US-09-548-373-3 Sequence 3, Appl1
9 1635.4 93.6 7983 96 US-60-184-771-693 Sequence 693, App
10 1629 93.2 2341 35 US-09-009-191-1 Sequence 1, Appl1
11 1611.4 92.2 2370 35 US-09-009-191-3 Sequence 3, Appl1
12 1489.6 85.3 1977 57 US-09-548-365-5 Sequence 5, Appl1
13 1489.6 85.3 1977 57 US-09-548-366-5 Sequence 5, Appl1
14 1489.6 85.3 1977 57 US-09-548-367-5 Sequence 5, Appl1
15 1489.6 85.3 1977 57 US-09-548-368-5 Sequence 5, Appl1
16 1489.6 85.3 1977 57 US-09-548-372-5 Sequence 5, Appl1
17 1489.6 85.3 1977 57 US-09-548-373-5 Sequence 5, Appl1
18 1473.8 84.4 1816 56 US-09-471-275-5645 Sequence 5645, Ap
19 1357.6 77.7 7509 92 US-09-172-360-26327 Sequence 26327, A
20 1242.6 71.1 2043 57 US-09-548-365-7 Sequence 7, Appl1
21 1242.6 71.1 2043 57 US-09-548-366-7 Sequence 7, Appl1
22 1242.6 71.1 2043 57 US-09-548-367-7 Sequence 7, Appl1
23 1242.6 71.1 2043 57 US-09-548-368-7 Sequence 7, Appl1
24 1242.6 71.1 2043 57 US-09-548-372-7 Sequence 7, Appl1
25 1242.6 71.1 2043 57 US-09-548-373-7 Sequence 7, Appl1
26 991.6 56.8 1278 57 US-09-548-365-27 Sequence 27, Appl1
27 991.6 56.8 1278 57 US-09-548-366-27 Sequence 27, Appl1
28 991.6 56.8 1278 57 US-09-548-367-27 Sequence 27, Appl1
29 991.6 56.8 1278 57 US-09-548-368-27 Sequence 27, Appl1
30 991.6 56.8 1278 57 US-09-548-372-27 Sequence 27, Appl1
31 991.6 56.8 1278 57 US-09-548-373-27 Sequence 27, Appl1
32 991.6 56.8 1302 57 US-09-548-365-25 Sequence 25, Appl1
33 991.6 56.8 1302 57 US-09-548-366-25 Sequence 25, Appl1
34 991.6 56.8 1302 57 US-09-548-367-25 Sequence 25, Appl1
35 991.6 56.8 1302 57 US-09-548-368-25 Sequence 25, Appl1
36 991.6 56.8 1302 57 US-09-548-372-25 Sequence 25, Appl1
37 991.6 56.8 1302 57 US-09-548-373-25 Sequence 25, Appl1
38 991.6 56.8 1341 57 US-09-548-365-21 Sequence 21, Appl1
39 991.6 56.8 1341 57 US-09-548-366-21 Sequence 21, Appl1
40 991.6 56.8 1341 57 US-09-548-367-21 Sequence 21, Appl1
41 991.6 56.8 1341 57 US-09-548-368-21 Sequence 21, Appl1
42 991.6 56.8 1341 57 US-09-548-372-21 Sequence 21, Appl1
43 991.6 56.8 1341 57 US-09-548-373-21 Sequence 21, Appl1
44 991.6 56.8 1362 57 US-09-548-365-29 Sequence 29, Appl1
45 991.6 56.8 1362 57 US-09-548-366-29 Sequence 29, Appl1

ALIGNMENTS

RESULT 1
US-09-215-435-66
; Sequence 66, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric;
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended CDNs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER FILING DATE: 1997-12-17
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 66
; LENGTH: 1747
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

; NAME/KEY: CDS
; LOCATION: 10..1062
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 10..57
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.9
; OTHER INFORMATION: seq FVYLAHFHTLCSG/WS
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1710..1715
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1735..1747
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1011
; OTHER INFORMATION: n=a, g, c or t
; US-09-215-435-66

Query Match 99.9%; Score 1746; DB 43; Length 1747;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gctcaccatagggtcccttcacatctctgcaagccacatttacactctgttctgggtg 60
|||||
Db 1 gctcaccatagggtcccttcacatctctgcaagccacatttacactctgttctgggtg 60
|||||
QY 61 tcagcacatccgggacacctccggaagggtgtgtatgtgcctacacccagggaagtgg 120
|||||
Db 61 tcagcacatccgggacacctccggaagggtgtgtatgtgcctacacccagggaagtgg 120
|||||
QY 121 gaaggagctgggacccagacctggtaagcatcccccatggccccacgtcaactgtcgt 180
|||||
Db 121 gaaggagctgggacccagacctggtaagcatcccccatggccccacgtcaactgtcgt 180
|||||
QY 181 gccaatctgtcctcactcaatgacagacaagtcttctatcaacgctccactgggaa 240
|||||
Db 181 gccaatctgtcctcactcaatgacagacaagtcttctatcaacgctccactgggaa 240
|||||
QY 241 ggcactctggggctgacctatgctgagattgccagacctcccgagccttc 300
Db 241 ggcactctggggctgacctatgctgagattgccagacctcccgagccttc 300
|||||
QY 301 ttgactctctgtaagcagacccacgttcccaacctcttccctgagctttgtggt 360
|||||
Db 301 ttgactctctgtaagcagacccacgttcccaacctcttccctgagctttgtggt 360
|||||
QY 361 gctggcttccctcctcaaccagctctgaagtgtgctgctctgtcgaggagcatgcatc 420
Db 361 gctggcttccctcctcaaccagctctgaagtgtgctgctctgtcgaggagcatgcatc 420
|||||
QY 421 gaggtatcgacacactcgtcttacacaggcagctctgtgtatataccatcccgccggag 480
Db 421 gaggtatcgacacactcgtcttacacaggcagctctgtgtatataccatcccgccggag 480
|||||
QY 481 tgggtattatgagtgatcattgtgcgggtggagatcaatggacagagatctgaaaaaggac 540
Db 481 tgggtattatgagtgatcattgtgcgggtggagatcaatggacagagatctgaaaaaggac 540
|||||
QY 541 tgcaaggagtacaactatgacaagagcattgtggcagtgggcaccacacacctctgttg 600
Db 541 tgcaaggagtacaactatgacaagagcattgtggcagtgggcaccacacacctctgttg 600
|||||
QY 601 cccaagaaagtgttgaagctgcagtcacaaatccatcaagcagcctcctccacggagaag 660
Db 601 cccaagaaagtgttgaagctgcagtcacaaatccatcaagcagcctcctccacggagaag 660
|||||
QY 661 ttccctgacggtttctgctaggagagcgtgtgtgtggaagcagcaccacct 720
Db 661 ttccctgacggtttctgctaggagagcgtgtgtgtggaagcagcaccacct 720
|||||

```
Qy 721 tggaaattttccagtcactctacttaataatgggtgaggttaccacacagtccttc 780
Db 721 tggaaattttccagtcactctacttaataatgggtgaggttaccacacagtccttc 780
Qy 781 cgeataccatcttcgcagcaataactcgcgcagtggaagatggccacagtcacca 840
Db 781 cgeataccatcttcgcagcaataactcgcgcagtggaagatggccacagtcacca 840
Qy 841 gacgactgttacagtttgccatctcacagtcatccacgagcactgttatgggagctgtt 900
Db 841 gacgactgttacagtttgccatctcacagtcatccacgagcactgttatgggagctgtt 900
Qy 901 atcatgagggcttcactgttctgtcttgatcgggcccgaacaaattgcttctgtctgc 960
Db 901 atcatgagggcttcactgttctgtcttgatcgggcccgaacaaattgcttctgtctgc 960
Qy 961 agcgttgcacatgtgcacgatgattcaggacgagcagcgttggaagccenttttgcac 1020
Db 961 agcgttgcacatgtgcacgatgattcaggacgagcagcgttggaagccenttttgcac 1020
Qy 1021 ctgtgacatggaagactgtgctacaacattccacagacagatgagtcacccctcatgac 1080
Db 1021 ctgtgacatggaagactgtgctacaacattccacagacagatgagtcacccctcatgac 1080
Qy 1081 catagctatgtcatgtgctgcactctgcctctctcatgctgcactctgcctcatggt 1140
Db 1081 catagctatgtcatgtgctgcactctgcctctctcatgctgcactctgcctcatggt 1140
Qy 1141 gtgtcagtggcgtgcctcgcctcgcctgcgcacgacgatgactttgctgtgatgacat 1200
Db 1141 gtgtcagtggcgtgcctcgcctcgcctgcgcacgacgatgactttgctgtgatgacat 1200
Qy 1201 ctccctgctgaagtgaagagcccatgggcagaagataggattccctctggaccacact 1260
Db 1201 ctccctgctgaagtgaagagcccatgggcagaagataggattccctctggaccacact 1260
Qy 1261 ccgtggttcactttgttcacaaatagtagacacacagatggccactgtggccagacactc 1320
Db 1261 ccgtggttcactttgttcacaaatagtagacacacagatggccactgtggccagacactc 1320
Qy 1321 aggacctcccccaccccaaatccctctgccttgatggagaagaaaaggctggcaagg 1380
Db 1321 aggacctcccccaccccaaatccctctgccttgatggagaagaaaaggctggcaagg 1380
Qy 1381 tgggttcaggagactgtacactgttaggacagacagaagaagaagaagacactctgttg 1440
Db 1381 tgggttcaggagactgtacactgttaggacagacagaagaagaagaagacactctgttg 1440
Qy 1441 cgggaatactttgttcacctcaaatttaagtcgggaaattctgctgtgaaacttcag 1500
Db 1441 cgggaatactttgttcacctcaaatttaagtcgggaaattctgctgtgaaacttcag 1500
Qy 1501 ccctgaaacctttgtacacattctttaaatctccaaaccccaaatctctctttcttta 1560
Db 1501 ccctgaaacctttgtacacattctttaaatctccaaaccccaaatctctctttcttta 1560
Qy 1561 gtttcagaagtactggcatacacagcaggttaaccttggctgtgtccctgtggtacacctg 1620
Db 1561 gtttcagaagtactggcatacacagcaggttaaccttggctgtgtccctgtggtacacctg 1620
Qy 1621 gcagagaagagaccaaagctgtttccctgctggccaaagtctagtagagagatgacacag 1680
Db 1621 gcagagaagagaccaaagctgtttccctgctggccaaagtctagtagagagatgacacag 1680
Qy 1681 ttgtctatttcttttagagacagggactgtataaacagcctaactatggttgcaaaaaa 1740
Db 1681 ttgtctatttcttttagagacagggactgtataaacagcctaactatggttgcaaaaaa 1740
Qy 1741 aaaaaa 1747
Db 1741 aaaaaa 1747
```

```
RESULT 2
US-60-096-116-57
: Sequence 57, Application US/60096116
: GENERAL INFORMATION:
: APPLICANT: Edwards, Jean-Baptiste Dumas Milne
: APPLICANT: Duclert, Aymeric
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 501 West Broadway
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92101-3505
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Win95
: SOFTWARE: Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/60/096,116
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelien, Ned A.
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: GENSET.038PR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 57:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1773 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: DOUBLE
: TOPOLOGY: LINEAR
: MOLECULE TYPE: CDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo Sapiens
: TISSUE TYPE: Substantia nigra
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 10..57
: IDENTIFICATION METHOD: Von Heijne matrix
: OTHER INFORMATION: score 4.9
: OTHER INFORMATION: seq FYIQAHTLCSG/WS
: FEATURE:
: NAME/KEY: poly_a_signal
: LOCATION: 1712..1717
: FEATURE:
: NAME/KEY: poly_a
: LOCATION: 1736..1749
: IDENTIFICATION METHOD: blastn2
: FEATURE:
: NAME/KEY: est
: LOCATION: 1566..1735
: IDENTIFICATION METHOD: blastn2
: OTHER INFORMATION: identity 100
: OTHER INFORMATION: region 1..170
: OTHER INFORMATION: id R12815
: FEATURE:
: NAME/KEY: est
: LOCATION: complement(1569..1735)
: IDENTIFICATION METHOD: blastn2
: OTHER INFORMATION: identity 100
: OTHER INFORMATION: region 11..177
: OTHER INFORMATION: id AA665039
: FEATURE:
: NAME/KEY: est
: LOCATION: complement(1570..1735)
: IDENTIFICATION METHOD: blastn2
: OTHER INFORMATION: identity 100
```



```
Qy 181 gccaacattgctgcactcactgaatcagacaaagtcttccatcaacggctccaaactgggaa 240
Db 181 GCCAACATTGCTGCATCCTCACTGAATCAGACAAAGTTCTTTCATCAAGGGCTCCAACTGGGAA 240
Qy 241 ggcatactcggggctggcctatgctgagattgccaggcctgcagactccccggagccttcc 300
Db 241 GGCATCCTGGGGCTGGCCTATGCTGAGATTGCCAGGCCCTGACGACTCCCGGAGCCTTTC 300
Qy 301 tttaactctctggttaaacagacaccagcttccaaactcttccctcagactttgtgt 360
Db 301 TTGACTCTCTGGTAAAGACAGACCACCGTTCCCAACCTCTTCTCCCTCAGCTTTGTGGT 360
Qy 361 gctggcttccccctcaacccagctggaagtctggtcctctgctggaggagcagatgacatt 420
Db 361 GCTGGCTTCCCCCTCAACCAAGCTGGAAGTCTGTGGCTCTGTGCGAGGAGGAGCATGATCATT 420
Qy 421 ggaggatcgcaccactcgtgtcacacaggcgtctctgtgtatatacaccatccgcgggag 480
Db 421 GGAGGTATCGACCACCTCGCTGTACACAGGCACTCTCTGGTATACACCCATCCGGCGGAG 480
Qy 481 tggattatgaggtgatctgtcgggtgagatcaatgacagaggtatctgaaatggac 540
Db 481 TGGTATTATGAGGTGATCATTTGTGGGGTGGAGATCAATGGACAGGATCTGAAAATGGAC 540
Qy 541 tgcaggaggtacaactatgacaagagcattgtggacagtggcaccaccaacctctcgtttg 600
Db 541 TGCAAGGAGTACAACTATGACAAAGCATTTGTGGACAGTGGCACCCACCAACCTTCGTTTG 600
Qy 601 cccaagaaagtgttgaagctgcagtcataatccatcaaggcagcctctccacggagag 660
Db 601 CCCAAGAAAGTGTTTGAAGCTGCAGTCAATCCATCAAGGAGCAGCTCTCCACGGAGAAG 660
Qy 661 ttccctgacggtttctgctaggagagcagctggtgtctggtggaagcaggaacacccct 720
Db 661 TTCCTGACGGTTTCTGCTAGGAGAGCAGCTGGTGTCTGGCAAGCAGGACCAACCCCT 720
Qy 721 tggaaactttcccagtcactctacctaattgggtgaggttaccacacagtccttc 780
Db 721 TGGAACTTTCCCAGTCATCTCACTCTACCTAATGGGTGAGGTATACCAACAGTCCCTTC 780
Qy 781 cgcataccactctccgcagcaatactctgcggcagctggaagatgtggccacgtcccaa 840
Db 781 CGCATCACCATCTTCCGCAGCAATACCTGGCGGCAGTGGAAAGATGTGCCACGCTCCCAA 840
Qy 841 gacactgttcaagtctgccaatctcacagtcacatccacggcagcactgttatggagcatt 900
Db 841 GACGACTGTACAAGTTTGGCCATCTCACAGTCATCCACGGGCAGCTGTATGGGAGCTGT 900
Qy 901 atcatggagggttctacgttgtcttcttgatcgggcccgaaacgaatggctttgtctgc 960
Db 901 ATCATGGAGGGCTTCTAGCTTGKCTTTGATCGGGCCCCGAAACGAATTTGGCTTGTCTGTC 960
Qy 961 agcgc-ttgccatgtgcacgatgag-ttcaggagcgcagcgttggaaagcccnctttgtc 1018
Db 961 AGCGCTTTGCCATGTGCAGATGAGTTTCAGGACGCGCAGCGGTGGAAGGCC-TTTTGTC 1019
Qy 1019 acctggacatggaagactgtgctacaacattccacagacagatgagtcacacctcatg 1078
Db 1020 ACCTTGGACATGGAAGACTGTGGCTACAACTTCACACAGACAGATGAGTCAACCCCTCATG 1079
Qy 1079 accatgactatgtcgtgctgca-tctggccctcttcatgctgcccactctgctcat 1137
Db 1080 ACCATAGGCTATGTGCTGGTGGCCATTTCTGCCCTCTTCATGCTGCCACTGTGCTCAT 1139
Qy 1138 ggtgtgtcagtggtgctcctcgtcgtcgtgcggcagcagcatgatgactttgtgtatga 1197
Db 1140 GGTGTGTAGTGGCGCTGCCTCCGCTGCTCGCGCAGCAGCATGATGATTTTGTGTATGA 1199
Qy 1198 catctccctcgtcgaagttaggagggcccatggcagaagaatagggattccctcggaccaca 1257
Db 1200 CATCTCCCTGTGTAAGTGAAGGGCCCATGGGCAGAAAGATAGGGATTCCTCTGGACCACA 1259
Qy 1258 cctcgtggttcaactttggttcacaagttaggagacagatggcacctgtggccagagcac 1317
```

```
Db 1260 CCWMGTGTTTACCTTTGGTCACAAGTAGGAGACACAGATGSCACCTGTGGCCAGAGCAC 1319
Qy 1318 ctcaaggacctcccccacaccccaaatgctcctgctccttgatggagaagaaaggctggca 1377
Db 1320 CTCAGGACCTCCCCACCCACCAAAATGCTCTGCTTGTATGGAGAGGAAAAGGCTGGCA 1379
Qy 1378 aggtgggttccaggagctgtacctgtagagacagaaagagaagaaagaaagcactctgc 1437
Db 1380 AGTGGGTTCCAGGACCTGTACCTGTAGGAGACAGAAAGAGAAAGAAAGACACTCTGC 1439
Qy 1438 tggcgggaatactcttggctcacctcaaatgaagtcgggaaattctgctgcttgaactt 1497
Db 1440 TGGCGGGAATACTCTTGTGTACCTCAAAATTTAAGTCGGGAAATTTCTGCTGTGAACCTT 1499
Qy 1498 cagccctgaacctttgtcaccaattccttttaattcctcaaccccaagattcttcttttc 1557
Db 1500 CAGCCCTGAACCTTTGTCAACCATTCCTTTAAATTTCCAACCCAAAGTATTTCTTTTC 1559
Qy 1558 ttagtctcagaaagtactggcatcacacaggttaccttggcgtgtgctcctgtgtgtacc 1617
Db 1560 TTAGTTTCAGAAAGTACTGGCATCACAGGATTACCTTGGCGTGTGTCCTGTGTGTACC 1619
Qy 1618 ctggcagaagaagacccaagcttcttccctgctgctgccaagtccagtaggagagatgca 1677
Db 1620 CTGGCAGAGAAGAGACCAAGCTTTGTTCCCTGCTGSCCAAGTCACTAGGAGAGATGCA 1679
Qy 1678 cagttgtctattgttttagagacaggagctgtataaacaagcctcaactgtgtgcaaaa 1737
Db 1680 CAGTTTGTCTATTGCTTTTAGAGACAGGGACTGTATAAACAGCCTAACATTTGTGCAAAA 1739
Qy 1738 aaaaaaaaaa 1747
Db 1740 AAAAAAaaaa 1749
```

```
RESULT 3
US-09-548-365-3
; Sequence 3. Application US/09548365
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; FILE OF INVENTION: 28341/6280A
; CURRENT APPLICATION NUMBER: US/09/548,365
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2070
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-548-365-3
```

```
Query Match 94.1%; Score 1644.8; DB 57; Length 2070;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
Qy 60 gtccagcacataccgggacccctccgaaggtgtgtatgtgccctcacccagggcaagt 119
|||||
```

```
Db 351 gtccagcaataccggagacatccggaagggtgtgtatgtccctacacccacggcgaagtg 410
QY 120 ggaaggagctgggcaaccgacctggttaagcatcccccattggcccacacgctactgtgcg 179
Db 411 ggaaggagctgggcaaccgacctggttaagcatcccccattggcccacacgctactgtgcg 470
QY 180 tgcacaacttgcctgcaatacactgaatacagacaagtttcttcaatacagcgtcccaactggga 239
Db 471 tgcacaacttgcctgcaatacactgaatacagacaagtttcttcaatacagcgtcccaactggga 530
QY 240 aggcactctgggctggccttatctgagattgccaggcctgacgactccccggagccttt 299
Db 531 aggcactctgggctggccttatctgagattgccaggcctgacgactccccggagccttt 590
QY 300 ctttgactctctgtaaaagcagaccacgcttcccacactcttctccctgacgtttgtgg 359
Db 591 ctttgactctctgtaaaagcagaccacgcttcccacactcttctccctgacgtttgtgg 650
QY 360 tgcctgcttcccctcaaccagctctgaagtgtcgtggcctctgtcggggagagcatgatcat 419
Db 651 tgcctgcttcccctcaaccagctctgaagtgtcgtggcctctgtcggggagagcatgatcat 710
QY 420 tggaggtatcgaccactcgtgtacacagcagctctctggtatcaccccatccgcggga 479
Db 711 tggaggtatcgaccactcgtgtacacagcagctctctggtatcaccccatccgcggga 770
QY 480 gtggtattagagtgatcatcttgctgggtggagatcaatggacagatctgaaatgga 539
Db 771 gtggtattagagtgatcatcttgctgggtggagatcaatggacagatctgaaatgga 830
QY 540 ctgcaaggagtacaactatgacaagagcattgtggacagtggcaccaccaactctcgttt 599
Db 831 ctgcaaggagtacaactatgacaagagcattgtggacagtggcaccaccaactctcgttt 890
QY 600 gcccaagaaagtgttgaaagctcaactcaaatccatcaaggcagcctctccacggagaa 659
Db 891 gcccaagaaagtgttgaaagctcagtcacaaatcccaaaaggcgccctccacggagaa 950
QY 660 gtccctgacggtttctggctagagagcagctggtgtgctggcagcagcggcaccacccc 719
Db 951 gtccctgaggtttctggctagagagcagctggtgtgctggcagcagcggcaccacccc 1010
QY 720 ttggaacatttcccagtgcatctcactctacactaaatgggtgaggtttaccaaccagtcctt 779
Db 1011 ttggaacatttcccagtgcatctcactctacactaaatgggtgaggtttaccaaccagtcctt 1070
QY 780 ccgcataccatcttcccagacaataacctcgcgcagtcgagtggaagatgtgccaagtccca 839
Db 1071 ccgcataccatcttcccagacaataacctcgcgcagtcgagtggaagatgtgccaagtccca 1130
QY 840 agacgactgttacaagtgtgcccactctcacagtcataccacgggcagcgtttatggagcgtgt 899
Db 1131 agacgactgttacaagtgtgcccactctcacagtcataccacgggcagcgtttatggagcgtgt 1190
QY 900 tatcatggagggtcttcaagttgtcttcttgatcggcccgaaacgaatgtggtttgctgt 959
Db 1191 tatcatggagggtcttcaagttgtcttcttgatcggcccgaaacgaatgtggtttgctgt 1250
QY 960 cagcgtctaccatgtgcagatgaattcagagcgcagcagtggaagcgccttttgtca 1019
Db 1251 cagcgtctgccaagtgcagatgaattcagagcgcagcagtggaagcgccttttgtca 1309
QY 1020 ccttgacatgggaagactgtggctacaacattccacagacagatgagtcaaccctcatga 1079
Db 1310 ccttgacatgggaagactgtggctacaacattccacagacagatgagtcaaccctcatga 1369
QY 1080 ccatagcctatgtcatgggtgccatctgcgcctcttcatgtgccactctgcctcatgg 1139
Db 1370 ccatagcctatgtcatgggtgccatctgcgcctcttcatgtgccactctgcctcatgg 1429
QY 1140 tgtgtcagtgggcgtccctccctgctcctgcgcagcagcatgatacttctgtgatgaca 1199
Db 1430 tgtgtcagtgggcgtccctccctgctcctgcgcagcagcatgatacttctgtgatgaca 1489
```

```
QY 1200 tctccctgctgaagtgaaggagggcccatgggcagagaataggggattccccctggaccacacc 1259
Db 1490 tctccctgctgaagtgaaggagggcccatgggcagagaataggggattccccctggaccacacc 1549
QY 1260 tccgtggttcaactttggtcacaaagttagggagacacagatggcacctgtggccagagacact 1319
Db 1550 tccgtggttcaactttggtcacaaagttagggagacacagatggcacctgtggccagagacact 1609
QY 1320 caggacccctcccacccacccaaatgcctctgccttgatggagagaaaaggctggcaag 1379
Db 1610 caggacccctcccacccacccaaatgcctctgccttgatggagagaaaaggctggcaag 1669
QY 1380 gtgggttccaggactgtacctgttaggagacagaaaagagaagaagaagcactctgctg 1439
Db 1670 gtgggttccaggactgtacctgttagaagacagaaaagagaagaagaagcactctgctg 1729
QY 1440 gcgggaataactcttggtcacctcaaaatttaagtgcgggaaattctgctgtgaaacttca 1499
Db 1730 gcgggaataactcttggtcacctcaaaatttaagtgcgggaaattctgctgtgaaacttca 1789
QY 1500 gcctgaacctttgt-caccattcctttaaattctcaaccccaagtatctctttct 1558
Db 1790 gcctgaacctttgtccaccattcctttaaattctcaaccccaagtatctctttct 1849
QY 1559 tagtttcagaagtactggcatcacgcaggtttacatttgctgtgcttccctgtggtaccc 1618
Db 1850 tagtttcagaagtactggcatcacgcaggtttacatttgctgtgcttccctgtggtaccc 1909
QY 1619 tggcagagaagaagacaagctgttccctgctggccaaagtccagtaggagagatgcac 1678
Db 1910 tggcagagaagaagacaagctgttccctgctggccaaagtccagtaggagagatgcac 1969
QY 1679 agtttctattgctttagacagcaggtgtataacaagcctaacattgggtgcaaaaa 1738
Db 1970 agtttctattgctttagacagcaggtgtataacaagcctaacattgggtgcaaaaa 2029
```

RESULT 4

```
US-09-548-366-3
; Sequence 3, Application US/09548366
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrichson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; FILE REFERENCE: 2831/6280A
; CURRENT APPLICATION NUMBER: US/09/548,366
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2070
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-548-366-3
```

Query Match 94.1%; Score 1644.8; DB 57; Length 2070;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Qy 60 gtccagcacataccgggacccctccggaagggtgtgtatgtgcccctacaccccgaggcaagt 119
Db 351 gtccagcacataccgggacccctccggaagggtgtgtatgtgcccctacaccccgaggcaagt 410
Qy 120 ggaaggaggagctgggacccagacctggtaagcaatcccccatggcccccaacgtcaactgtgcg 179
Db 411 ggaaggaggagctgggacccagacctggtaagcaatcccccatggcccccaacgtcaactgtgcg 470
Qy 180 tgcacaatgtcctccatcactgaatcagacaaagtcttctcatcaacggctccaactggga 239
Db 471 tgcacaatgtcctccatcactgaatcagacaaagtcttctcatcaacggctccaactggga 530
Qy 240 aggcacctggggctggcctatgtgatgattgccaggcctgacgactcccgaggaccttt 299
Db 531 aggcacctggggctggcctatgtgatgattgccaggcctgacgactcccgaggaccttt 590
Qy 300 ctttgactctctgttaaagcagacccacgttcccacacctctctccctgcagctttgtgg 359
Db 591 ctttgactctctgttaaagcagacccacgttcccacacctctctccctgcacctttgtgg 650
Qy 360 tgcctggcttcccctcaaccagctgaagtgtgctgctgctctgtcgaggagcatgacat 419
Db 651 tgcctggcttcccctcaaccagctgaagtgtgctgctgctctgtcgaggagcatgacat 710
Qy 420 tggagggtatcgaccactcgtgttacacaggcagctctctgtgtatacaccctccggcgagg 479
Db 711 tggagggtatcgaccactcgtgttacacaggcagctctctgtgtatacaccctccggcgagg 770
Qy 480 gtggtattatgagtgatcatctgtgcgggtggagatcaatggacagagatctgaaatggga 539
Db 771 gtggtattatgagtgatcatctgtgcgggtggagatcaatggacagagatctgaaatggga 830
Qy 540 ctgcaaggagtacaciatgacaaagacattgtgacagtgagcaccaccaactctgttt 599
Db 831 ctgcaaggagtacaciatgacaaagacattgtgacagtgagcaccaccaactctgttt 890
Qy 600 gcccaagaaagtgttgaagctgcagtcacaaatccatcaagcagcctctccctccagcgagaa 659
Db 891 gcccaagaaagtgttgaagctgcagtcacaaatccatcaagcagcctctccctccagcgagaa 950
Qy 660 gtccctcgacggttctggctagagagcagctggtgtgctggtggaagcagggacaccccc 719
Db 951 gtccctcgatggtttctggctagagagcagctggtgtgctggtggaagcagggacaccccc 1010
Qy 720 ttgaaacattttccagtcactcatcactcaactaataagggtgagttaccacaacagctcct 779
Db 1011 ttgaaacattttccagtcactcatcactcaactaataagggtgagttaccacaacagctcct 1070
Qy 780 ccgcatcacatccttccgacgcaatacctcggccagtggaagatgtggccacgtccca 839
Db 1071 ccgcatcacatccttccgacgcaatacctcggccagtggaagatgtggccacgtccca 1130
Qy 840 agacgactgttacaagtttgccatctcacagtcacacgggacgactgttatggagctgt 899
Db 1131 agacgactgttacaagtttgccatctcacagtcacacgggacgactgttatggagctgt 1190
Qy 900 tatcatgagggcttctacgttgtctttgtatcgggcccgaaaaaagaaatggctttgtgt 959
Db 1191 tatcatgagggcttctacgttgtctttgtatcgggcccgaaaaaagaaatggctttgtgt 1250
Qy 960 cagcgttgcctatgcacgatgagttcagacgagcagcagtggaagggccctttttgtca 1019
Db 1251 cagcgttgcctatgcacgatgagttcagacgagcagcagtggaagggccctt-tttgtca 1309
Qy 1020 ccttggacatggaagactgtggttacaacattccacagacagatgagtcacacctcatga 1079
Db 1310 ccttggacatggaagactgtggttacaacattccacagacagatgagtcacacctcatga 1369
Qy 1080 ccatagcctatgtcatggtgccatctgcgcctcttcttcatgtgccactctgctcatgg 1139
Db 1370 ccatagcctatgtcatggtgccatctgcgcctcttcttcatgtgccactctgctcatgg 1429
Qy 1140 tgtgtcagtggcgtcctcgcgtgcctgcgcgcagcagcatgatgactttgtgatgaca 1199

Db 1430 tgtgtcagtggcgtcctcgcgtgcctgcgcgcagcagcatgatgactttgtgatgaca 1489
Qy 1200 tctccctgtcgaagtgaagaggcccatgggcagaagatagaggattcccttgagcacacc 1259
Db 1490 tctccctgtcgaagtgaagaggcccatgggcagaagatagaggattcccttgagcacacc 1549
Qy 1260 tccgtggtcacttctgtccaaagttaggagacacagatggcacctgtggccagagcacct 1319
Db 1550 tccgtggtcacttctgtccaaagttaggagacacagatggcacctgtggccagagcacct 1609
Qy 1320 caggaccttccccccacccacccaaatgcctcctgtccttgatgggagaagaaaaagctggcaag 1379
Db 1610 caggaccttccccccacccacccaaatgcctcctgtccttgatgggagaagaaaaagctggcaag 1669
Qy 1380 gtgggttccaggagactgtacctgttaggagacagaaaaagagaagaagacactctgctg 1439
Db 1670 gtgggttccaggagactgtacctgttaggaaacagaaaaagagaagaagacactctgctg 1729
Qy 1440 gcgggaatacactctgttcacctcaaatctaagtgcgggaaattctgctcttgaacttca 1499
Db 1730 gcgggaatacactctgttcacctcaaatctaagtgcgggaaattctgctcttgaacttca 1789
Qy 1500 gccctgaacctttgt-caccattccttttaaatctcccaaccccaaatgtattctttct 1558
Db 1790 gccctgaacctttgtccacattccttttaaatctcccaaccccaaatgtattctttct 1849
Qy 1559 tagttcagaagtagctggcatcacacgcaggttaaccttggcgtgtgctccctgtgtaccc 1618
Db 1850 tagttcagaagtagctggcatcacacgcaggttaaccttggcgtgtgctccctgtgtaccc 1909
Qy 1619 tggcagagaagacacacagctgtttccctctggtcccaaatcagtagagagatgacac 1678
Db 1910 tggcagagaagacacacagctgtttccctctggtcccaaatcagtagagagatgacac 1969
Qy 1679 agttgtctattgtcttttagagacagggagctgtataaaacacccctaaacttgggtcaaaaa 1738
Db 1970 agttgtctattgtcttttagagacagggagctgtataaaacacccctaaacttgggtcaaaaa 2029

RESULT 5

US-09-548-367-3
; Sequence 3, Application US/09548367
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; FILE REFERENCE: 28341/6280A
; CURRENT APPLICATION NUMBER: US/09/548,367
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2070
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-548-367-3

Query Match 94.1%; Score 1644.8; DB 57; Length 2070;
Best Local Similarity 99.5%; Pred. No. 0;

Matches 1671; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

```
QY 60 gtccagcacatccggaaccccggaagggtgtgtatgtccctacaccacagggcaagt 119
Db 351 gtccagcacatccggaaccccggaagggtgtgtatgtccctacaccacagggcaagt 410
QY 120 ggaagggaagctgggcaaccagactggttaagcatcccccattggcccacgtcaactgtcg 179
Db 411 ggaagggaagctgggcaaccagactggttaagcatcccccattggcccacgtcaactgtcg 470
QY 180 tgcacaattgtctgcatactgaatcagacaagttcttcatcaacggctcccaactggga 239
Db 471 tgcacaattgtctgcatactgaatcagacaagttcttcatcaacggctcccaactggga 530
QY 240 aggaactctgggctggcctatgtgagattgccaggcctgacgactcccccagaccttt 299
Db 531 aggaactctgggctggcctatgtgagattgccaggcctgacgactcccccagaccttt 590
QY 300 ctttgactctctggttaagcagaccacgcttcccacacttcttcctcgtcagctttgtgg 359
Db 591 ctttgactctctggttaagcagaccacgcttcccacacttcttcctcgtcagctttgtgg 650
QY 360 tgcgtggttccctcccaaccagctgaagtgtgctggtgctctgtcgaggagcagcatgat 419
Db 651 tgcgtggttccctcccaaccagctgaagtgtgctggtgctctgtcgaggagcagcatgat 710
QY 420 tggaggtatcgaccactcgtgtacacagcagctctctgtgtatcacccatccggcgga 479
Db 711 tggaggtatcgaccactcgtgtacacagcagctctctgtgtatcacccatccggcgga 770
QY 480 gtggtattagaggtatcatattgtggtggtgagatcaatggacaggaatgtgaaatgga 539
Db 771 gtggtattagaggtatcatattgtggtggtgagatcaatggacaggaatgtgaaatgga 830
QY 540 ctgcaaggagtaacaactatgacaagagcatgtggacagtggcaccaccaactcgttt 599
Db 831 ctgcaaggagtaacaactatgacaagagcatgtggacagtggcaccaccaactcgttt 890
QY 600 gcccaagaaagtgttgaagctgcagtcacaaatccatcaaggcagcctctccacggagaa 659
Db 891 gcccaagaaagtgttgaagctgcagtcacaaatccatcaaggcagcctctccacggagaa 950
QY 660 gtccctcagcgtttctggttagagagagcagctggtgtgctggcaaggcagccaccccc 719
Db 951 gtccctcagcgtttctggttagagagagcagctggtgtgctggcaaggcagccaccccc 1010
QY 720 ttggaacattttcccgactcactcactcaatgggtgaggttaccacacagctcctt 779
Db 1011 ttggaacattttcccgactcactcactcaatgggtgaggttaccacacagctcctt 1070
QY 780 ccgcataccatcttcgcagcaatctcctggcgccagtggaagatgtggccacgtccca 839
Db 1071 ccgcataccatcttcgcagcaatctcctggcgccagtggaagatgtggccacgtccca 1130
QY 840 agacgactgttaacatttgccatctcacagtcacatccacgggcactgttatgggagctgt 899
Db 1131 agacgactgttaacatttgccatctcacagtcacatccacgggcactgttatgggagctgt 1190
QY 900 tatcatggagggtcttctacgttgtcttctgatcgggcccgaaacgaattggcttgcgt 959
Db 1191 tatcatggagggtcttctacgttgtcttctgatcgggcccgaaacgaattggcttgcgt 1250
QY 960 cagcgttccatgtgcacatgagttcagagcagcggtggagggccncttttgtca 1019
Db 1251 cagcgttccatgtgcacatgagttcagagcagcggtggagggcc-cttttgtca 1309
QY 1020 ccttgacatggaagactgtgctacaacattccacagacagatgagtcacccctcatga 1079
Db 1310 ccttgacatggaagactgtgctacaacattccacagacagatgagtcacccctcatga 1369
QY 1080 ccatagcctatgtcatgggtgcacatgcccctcttctatgctgcaactctgcctcatgg 1139
Db 1370 ccatagcctatgtcatgggtgcacatgcccctcttctatgctgcaactctgcctcatgg 1429
```

```
QY 1140 tgtgtcagtgagcgtgctccgctgctgcgcagcagcatgatgactttgtcgtgatgaca 1199
Db 1430 tgtgtcagtgagcgtgctccgctgctgcgcagcagcatgatgactttgtcgtgatgaca 1489
QY 1200 tctccctgtcgaagtgtgagagggcccatgggcagagaatagggatccctcgtggaccaccc 1259
Db 1490 tctccctgtcgaagtgtgagagggcccatgggcagagaatagagatccctcgtggaccaccc 1549
QY 1260 tccggtgttcacattgttcacaagttaggagacacatggcacctgtggccagagacact 1319
Db 1550 tccggtgttcacattgttcacaagttaggagacacatggcacctgtggccagagacact 1509
QY 1320 caggacccctcccccacccacccaaatgctctgcttgcagaggaagaaaggctggcaag 1379
Db 1610 caggacccctcccccacccacccaaatgctctgcttgcagaggaagaaaggctggcaag 1669
QY 1380 gtgggttcaggagactgtaccctgttaggagacagaaagaagaagaagcactctgctg 1439
Db 1670 gtgggttcaggagactgtaccctgttaggagacagaaagaagaagaagcactctgctg 1729
QY 1440 gcgggaatactctgtgtcacctcacaatttaagtcgggaattctgctgttgaacttca 1499
Db 1730 gcgggaatactctgtgtcacctcacaatttaagtcgggaattctgctgttgaacttca 1789
QY 1500 gcctgaacctttgt-caccattccttttaattctccaccccaaaagtattctctttct 1558
Db 1790 gcctgaacctttgttcaccattccttttaattctccaccccaaaagtattctctttct 1849
QY 1559 tagtttcagaagtactgtgcatcacacgcaggtttacacttggtgtcctcgttgggtacc 1618
Db 1850 tagtttcagaagtactgtgcatcacacgcaggtttacacttggtgtcctcgttgggtacc 1909
QY 1619 tgccagagaagacacaaagctgtttccctgctgcccgaagtcagtaggagagatgcac 1678
Db 1910 tgccagagaagacacaaagctgtttccctgctgcccgaagtcagtaggagagatgcac 1969
QY 1679 agttgtctattgtcttagacagggactgtataaacagcctaacattggtgcacaaa 1738
Db 1970 agttgtctattgtcttagacagggactgtataaacagcctaacattggtgcacaaa 2029
```

```
RESULT 6
US-09-548-368-3
; Sequence 3, Application US/09548368
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES THEREFOR
; FILE REFERENCE: 28341/6280A
; CURRENT APPLICATION NUMBER: US/09/548,368
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2070
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-548-368-3
```


Query Match		94.1%;	Score 1644.8;	DB 57;	Length 2070;
Best Local Similarity		99.5%;	Pred. No. 0;		
Matches 1671;		Conservative	0;	Mismatches	7;
				Indels	2;
				Gaps	2;
Qy	60	gtccagcacataccgggacctcccgaaagggtgtgtatgtgcccctacaccccgaggcaagt	119		
Db	351	gtccagcacataccgggacctcccgaaagggtgtgtatgtgcccctacaccccgaggcaagt	410		
Qy	120	ggaaggaggagtgggcaccgacctggtaagcatccccatgcccacacgtcactgtgcg	179		
Db	411	ggaaggaggagtgggcaccgacctggtaagcatccccatgcccacacgtcactgtgcg	470		
Qy	180	tgcaaacattgtccatcaactgaatcacagacaagtctctcatcacggctccaaactggga	239		
Db	471	tgcaaacattgtccatcaactgaatcacagacaagtctctcatcacggctccaaactggga	530		
Qy	240	aggaacctggggctgacctatgtgtgagattgccaggcctgacgactccccggagccttt	299		
Db	531	aggaacctggggctgacctatgtgtgagattgccaggcctgacgactccccggagccttt	590		
Qy	300	ctttgactctctgttaaagcagacccacgttcccaacctctctccctgcaagctttgtg	359		
Db	591	ctttgactctctgttaaagcagacccacgttcccaacctctctccctgcaagctttgtg	650		
Qy	360	tgctggcttccccctcaaccagctggaagtgtgctgctctgtcgaggaggagcatgatcat	419		
Db	651	tgctggcttccccctcaaccagctggaagtgtgctgctctgtcgaggaggagcatgatcat	710		
Qy	420	tggagggtatcgaccactcgtgtcacaggcagtcctctgtgtatcacaccatccggcgggga	479		
Db	711	tggagggtatcgaccactcgtgtcacaggcagtcctctgtgtatcacaccatccggcgggga	770		
Qy	480	gtggtattatgagtgatcatgtgcgggtgagagataaatgagacagatctgaataatgga	539		
Db	771	gtggtattatgagtgatcatgtgcgggtgagagataaatgagacagatctgaataatgga	830		
Qy	540	ctgcaaggagtacaactatgacaagagcattgtgacagtgggcaccacacacctcgctt	599		
Db	831	ctgcaaggagtacaactatgacaagagcattgtgacagtgggcaccacacacctcgctt	890		
Qy	600	gcccagaagaagtgtttgaagctgcagtcacaaatccatcaagcagcctctccacggagaa	659		
Db	891	gcccagaagaagtgtttgaagctgcagtcacaaatccatcaagcagcctctccacggagaa	950		
Qy	660	gttccctgacgggtttctggcttaggagagacagctgggtgtctggtggaagcaggcacacc	719		
Db	951	gttccctgacgggtttctggcttaggagagacagctgggtgtctggtggaagcaggcacacc	1010		
Qy	720	tggacaattttccagtcactctcaactacctaagtgggtgaggttaccacacagtcctt	779		
Db	1011	tggacaattttccagtcactctcaactacctaagtgggtgaggttaccacacagtcctt	1070		
Qy	780	ccgcatcacctctcccgacgacataacctgggccagtggaagatgtggccacgtccca	839		
Db	1071	ccgcatcacctctcccgacgacataacctgggccagtggaagatgtggccacgtccca	1130		
Qy	840	agacgactgttacaaagtgtgcatctcacatcagtcacacggcagcagctgttatggagctgt	899		
Db	1131	agacgactgttacaaagtgtgcatctcacatcagtcacacggcagcagctgttatggagctgt	1190		
Qy	900	tatcatgaggggcttctacgttctgttcttgatcgggcccgaaacgaattggctttgctgt	959		
Db	1191	tatcatgaggggcttctacgttctgttcttgatcgggcccgaaacgaattggctttgctgt	1250		
Qy	960	cagcgttggcatgtgacgatgagttcaggacggcagcagcgggtggaagcccnntttgtca	1019		
Db	1251	cagcgttggcatgtgacgatgagttcaggacggcagcagcgggtggaagcccnntttgtca	1309		
Qy	1020	ctttggacatggaagactgtgtgtacaaacttccacagacagatgagtcacacctcatga	1079		
Db	1310	ctttggacatggaagactgtgtgtacaaacttccacagacagatgagtcacacctcatga	1369		
Qy	1080	ccatagcctatgtcatggtgccatctgcgcctcttctcatgctgccactctgctcatcgtg	1139		

Db	1370	ccatagcctatgtcatggtgccatctgcgcctcttctcatgctgccactctgctcatcgtg	1429		
Qy	1140	tgtgtcagtggcgtgcctcgcgtgcctgcgcagcagcatgatgacttttgcgtgatgaca	1199		
Db	1430	tgtgtcagtggcgtgcctcgcgtgcctgcgcagcagcatgatgacttttgcgtgatgaca	1489		
Qy	1200	tctccctgtcaagttaggagggcccatggtgcagaagatagggattccctcctgaccacacc	1259		
Db	1490	tctccctgtcaagttaggagggcccatggtgcagaagatagggattccctcctgaccacacc	1549		
Qy	1260	tccgtgggttcactttgttcacaagttaggacacagatggcacctgtggccagagcacct	1319		
Db	1550	tccgtgggttcactttgttcacaagttaggacacagatggcacctgtggccagagcacct	1609		
Qy	1320	cagacacctccccaccacccaaatgcctctgcctctgtatggagaaggaaaaagcgtgcgaag	1379		
Db	1610	cagacacctccccaccacccaaatgcctctgcctctgtatggagaaggaaaaagcgtgcgaag	1669		
Qy	1380	gtgggttccaggagctgtacctgttaggagacagaaaaagagaagaagcactctgctg	1439		
Db	1670	gtgggttccaggagctgtacctgttaggaaacagaaaaagagaagaagcactctgctg	1729		
Qy	1440	gcgggaatactctgtgcacctcaaatttaagtcgggaaattctgctgttgaaacttca	1499		
Db	1730	gcgggaatactctgtgcacctcaaatttaagtcgggaaattctgctgttgaaacttca	1789		
Qy	1500	gcctgaaacctttgt-caccattctttaaattctccaccccaaaagtattctctttct	1558		
Db	1790	gcctgaaacctttgtccaccattctttaaattctccaccccaaaagtattctctttct	1849		
Qy	1559	tagtttcagaagtactggcaccacagcaggttaaccttggcgtgtgctcctgtgtaccc	1618		
Db	1850	tagtttcagaagtactggcaccacagcaggttaaccttggcgtgtgctcctgtgtaccc	1909		
Qy	1619	tggcagagaagagaccaaagctgtttccctgcgtgcccacaaagtcagtaggaggtgcac	1678		
Db	1910	tggcagagaagagaccaaagctgtttccctgcgtgcccacaaagtcagtaggaggtgcac	1969		
Qy	1679	agttgtctattgtctttagagacaggagctgtataaaacagcctaacattggtgcacaaa	1738		
Db	1970	agttgtctattgtctttagagacaggagctgtataaaacagcctaacattggtgcacaaa	2029		

RESULT 7
US-09-548-372-3
; Sequence 3, Application US/09548372
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riliang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES THEREFOR
; FILE REFERENCE: 28341/6280A
; CURRENT APPLICATION NUMBER: US/09/548.372
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 2070
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-548-372-3

Query Match									
Best Local Similarity			94.1%; Score 1644.8; DB 57; Length 2070;		Pred. No. 0;				
Matches 1671; Conservative			0; Mismatches		7; Indels		2; Gaps		
Qy	60	gtccagcacatccggagactccggaaggtgtgtatgtccctacacccaggccaagt	119						
Db	351	gtccagcacatccggagactccggaaggtgtgtatgtccctacacccaggccaagt	410						
Qy	120	ggaaggaggctgggaccgagactggtaagcatcccatggcccaacgtcactgtgcg	179						
Db	411	ggaaggaggctgggaccgagactggtaagcatcccatggcccaacgtcactgtgcg	470						
Qy	180	tgccacaattgtgcatacactgaatcagaagaattcttcatacaacgctccaactggga	239						
Db	471	tgccacaattgtgcatacactgaatcagaagaattcttcatacaacgctccaactggga	530						
Qy	240	aggcatctgggctggcctatgtcgagattgccaggctgaacactcccgagactctt	299						
Db	531	aggcatctgggctggcctatgtcgagattgccaggctgaacactcccgagactctt	590						
Qy	300	ctttgactctctgtgtaagcagaccgaattccccaaactcttccttcgagcttttgg	359						
Db	591	ctttgactctctgtgtaagcagaccgaattccccaaactcttccttcgagcttttgg	650						
Qy	360	tgctggcttcccccaaccagctctgaagtctggcctctatcgaggagacatgacat	419						
Db	651	tgctggcttcccccaaccagctctgaagtctggcctctatcgaggagacatgacat	710						
Qy	420	tggaggtatcgacactcgtgtacacaggcagctctcgtgtatacacccatccgcggga	479						
Db	711	tggaggtatcgacactcgtgtacacaggcagctctcgtgtatacacccatccgcggga	770						
Qy	480	gtggtattatgaagtgaatcatgtgcgggtggagatacaatggacagatctgaaaaatga	539						
Db	771	gtggtattatgaagtgaatcatgtgcgggtggagatacaatggacagatctgaaaaatga	830						
Qy	540	ctgcaaggagtacaactatgacaagagcattgtgacagtgggcaccaccaactctgtt	599						
Db	831	ctgcaaggagtacaactatgacaagagcattgtgacagtgggcaccaccaactctgtt	890						
Qy	600	gcccaagaagaagtgttgaaagctgcagtcacaaatccataaggcagcctcctccacggaa	659						
Db	891	gcccaagaagaagtgttgaaagctgcagtcacaaatccataaggcagcctcctccacggaa	950						
Qy	660	gttccctgacgggttcttgctgtaggagagcagctggtgtgctggcaagcaggcacacccc	719						
Db	951	gttccctgacgggttcttgctgtaggagagcagctggtgtgctggcaagcaggcacacccc	1010						
Qy	720	tgggaacattttcccagtcatactcacttaactaaatgggtgaggtaccacaaccagctctt	779						
Db	1011	tgggaacattttcccagtcatactcacttaactaaatgggtgaggtaccacaaccagctctt	1070						
Qy	780	ccgcataccactctccgcagcaataactcggccagtgaaagatgtgccacgtccca	839						
Db	1071	ccgcataccactctccgcagcaataactcggccagtgaaagatgtgccacgtccca	1130						
Qy	840	agacgactgttaacaagtttggcatctccagtcataccacgggcactgttatgggagctgt	899						
Db	1131	agacgactgttaacaagtttggcatctccagtcataccacgggcactgttatgggagctgt	1190						
Qy	900	tatcatggagggtcttaagttgttattgataggcccgaaaaacgaatttggctttgctgt	959						
Db	1191	tatcatggagggtcttaagttgttattgataggcccgaaaaacgaatttggctttgctgt	1250						
Qy	960	cagcgcttgcattgtcacagatgagttcagacgcagcgttgaaaggcccttttgtca	1019						
Db	1251	cagcgcttgcattgtcacagatgagttcagacgcagcgttgaaaggcccttttgtca	1309						
Qy	1020	ccttgacatgaagactgtggctacaactt							

RESULT 8
US-09-548-373-3
Sequence 3, Application US/09548373
GENERAL INFORMATION:
APPLICANT: Gurney, Mark E.
APPLICANT: Bienkowski, Michael J.
APPLICANT: Heinrikson, Robert L.
APPLICANT: Parodi, Luis A.
APPLICANT: Yan, Riqiang
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
TITLE OF INVENTION: USES THEREFOR
FILE REFERENCE: 28341/6280A
CURRENT APPLICATION NUMBER: US/09/548, 373
CURRENT FILING DATE: 2000-04-12
PRIORITY APPLICATION NUMBER: 60/155,493
PRIORITY FILING DATE: 1999-09-23
PRIORITY APPLICATION NUMBER: 09/404,133
PRIORITY FILING DATE: 1999-09-23
PRIORITY APPLICATION NUMBER: PCT/US99/20881
PRIORITY FILING DATE: 1999-09-23
PRIORITY APPLICATION NUMBER: 60/101,594
PRIORITY FILING DATE: 1998-09-24
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2070
TYPE: DNA

ORGANISM: Homo sapiens
US-09-548-373-3

Query Match 94.1%; Score 1644.8; DB 57; Length 2070;
Best Local Similarity 99.5%; Pred. NO. 0;
Matches 1671; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 60 gtcacgacatccgggacccctcgaaaggtgtgtatgtgccctacacccaggccaagtg 119
DB 351 gtcacgacatccgggacccctcgaaaggtgtgtatgtgccctacacccaggccaagtg 410

QY 120 ggaaggagctgggacacgacctgtaagcatcccccacgacgacacgacgacgacg 179
DB 411 ggaaggagctgggacacgacctgtaagcatcccccacgacgacgacgacgacg 470

QY 180 tgcacacattgctgcatacactgaatcagacagaagtcttcatacagcgtccaaatggga 239
DB 471 tgcacacattgctgcatacactgaatcagacagaagtcttcatacagcgtccaaatggga 530

QY 240 aggcacccctgggctggcctatgctgagattgccaggcctgacgactcccgagccctt 299
DB 531 aggcacccctgggctggcctatgctgagattgccaggcctgacgactcccgagccctt 590

QY 300 ctttgactctctggtaaagcagaccacgcttcccaacctctctccctgcagctttgtg 359
DB 591 ctttgactctctggtaaagcagaccacgcttcccaacctctctccctgcacctttgtg 650

QY 360 tgcctggctcccccacacagctgtaagtgctggcctctgtcggaggagcatgatcat 419
DB 651 tgcctggctcccccacacagctgtaagtgctggcctctgtcggaggagcatgatcat 710

QY 420 tgaaggtatcgaccactgctgtacacaggcagctctctgtgtatatacccatccggcgga 479
DB 711 tgaaggtatcgaccactgctgtacacaggcagctctctgtgtatatacccatccggcgga 770

QY 480 ggggtattatgagtgatcatgtgcgggtggagatcaatgacagagatctgaataatga 539
DB 771 ggggtattatgagtgatcatgtgcgggtggagatcaatgacagagatctgaataatga 830

QY 540 ctgcaaggagtagcaactatgacaagagcatgtggacagtggcacacacacccctcgctt 599
DB 831 ctgcaaggagtagcaactatgacaagagcatgtggacagtggcacacacacccctcgctt 890

QY 600 gcccaagaagtgtttgaagctgcagtcacaaatccatcaaggcagcctctccacaggagaa 659
DB 891 gcccaagaagtgtttgaagctgcagtcacaaatccatcaaggcagcctctccacaggagaa 950

QY 660 gttccctgacggtttctggctaggagagcagctggtgtgctggcaagcaggcaccacccc 719
DB 951 gttccctgaggtttctggctaggagagcagctggtgtgctggcaagcaggcaccacccc 1010

QY 720 ttgggaacattttcccgatcatctcactacataatgggtgaggttaccaaccagtcctt 779
DB 1011 ttgggaacattttcccgatcatctcactacataatgggtgaggttaccaaccagtcctt 1070

QY 780 ccgcatacaccatcttcgcagcaatacctgcggccagtggaagatgtggccacgtccca 839
DB 1071 ccgcatacaccatcttcgcagcaatacctgcggccagtggaagatgtggccacgtccca 1130

QY 840 agacgactgtacaagtgtgcatctccagtcacagtcacccggcagcactgttatggagctgt 899
DB 1131 agacgactgtacaagtgtgcatctccagtcacagtcacccggcagcactgttatggagctgt 1190

QY 900 tatcatggagggtctctacgtgtctttgatcgggcccgaacaaacgaattggtttgctgt 959
DB 1191 tatcatggagggtctctacgtgtctttgatcgggcccgaacaaacgaattggtttgctgt 1250

QY 960 caggccttgccatgtgcagatgagtccaggcagcagcggtggaaggcccttttgtca 1019
DB 1251 caggccttgccatgtgcagatgagtccaggcagcagcggtggaaggcccttttgtca 1309

QY 1020 ccttggacatggaagactgtggtacataacattccacagacagatgagtcacccctcatga 1079

DB 1310 ccttggacatggaagactgtggtacataattccacagacagatgagtcacccctcatga 1369

QY 1080 ccatagcctatgtatgctgcatctggtgcctcttcatgtgcacactctgcctcatgg 1139
DB 1370 ccatagcctatgtatgctgcatctggtgcctcttcatgtgcacactctgcctcatgg 1429

QY 1140 tgtgtcagtggtgcctgcctccctgcctgcgcacgacgacatgatgactttgctgatgaca 1199
DB 1430 tgtgtcagtggtgcctgcctccctgcctgcgcacgacgacatgatgactttgctgatgaca 1489

QY 1200 tctccctgtgaagtggaggagggcccatgggcagaagatagggattccccctggaccacacc 1259
DB 1490 tctccctgtgaagtggaggagggcccatgggcagaagatagagattccccctggaccacacc 1549

QY 1260 tccgtgtgttcatttggtcacaagttaggagacacagatggcaccctgtgcccagagcacct 1319
DB 1550 tccgtgtgttcatttggtcacaagttaggagacacagatggcaccctgtgcccagagcacct 1609

QY 1320 caggacctccccaccaccacaaatgctctgctgtgatggagaagaaagcgtggcaag 1379
DB 1610 caggacctccccaccaccacaaatgctctgctgtgatggagaagaaagcgtggcaag 1669

QY 1380 gtgggttcaggagactgtacctgtaggagacagaaaaagagaagaagacactctgctg 1439
DB 1670 gtgggttcaggagactgtacctgtaggagacagaaaaagagaagaagacactctgctg 1729

QY 1440 ggggaataactcttggtcacacctcaaatatttaagtcgggaaattctgtctgtgaaactca 1499
DB 1730 ggggaataactcttggtcacacctcaaatatttaagtcgggaaattctgtctgtgaaactca 1789

QY 1500 gccctgaacctttgt-caccattctttaattctcaaccccaaaagtattctttctt 1558
DB 1790 gccctgaacctttgtccaccattctttaattctcaaccccaaaagtattctttctt 1849

QY 1559 tagtttcagaagtactggcatcacacgcagggttacccttggctgtgctccctgtggtacc 1618
DB 1850 tagtttcagaagtactggcatcacacgcagggttacccttggctgtgctccctgtggtacc 1909

QY 1619 tggcagagaagagaccaagctgtttccctgctggtcccaagtcagtaggagaggtgcac 1678
DB 1910 tggcagagaagagaccaagctgtttccctgctggtcccaagtcagtaggagaggtgcac 1969

QY 1679 agttgtctatttgccttttagacagagagctgtataaacagcctaactgttgcaaaaa 1738
DB 1970 agttgtctatttgccttttagacagagagctgtataaacagcctaactgttgcaaaaa 2029

RESULT 9

US-60-184-771-693

; Sequence 693, Application US/60184771

; GENERAL INFORMATION:

; APPLICANT: Hodgson, David M.

; APPLICANT: Lincoln, Stephen E.

; APPLICANT: Russo, Frank D.

; APPLICANT: Spiro, Peter A.

; APPLICANT: Banville, Steve C.

; APPLICANT: Brufcher, Shawn R.

; APPLICANT: Dufour, Gerard E.

; APPLICANT: Cohen, Howard J.

; APPLICANT: Rosen, Bruce

; APPLICANT: Shah, Purvi

; APPLICANT: Chalup, Michael S.

; APPLICANT: Hillman, Jennifer L.

; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES

; FILE REFERENCE: PT-0120 P

; CURRENT APPLICATION NUMBER: US/60/184,771

; CURRENT FILING DATE: 2000-02-24

; NUMBER OF SEQ ID NOS: 1244

; SOFTWARE: PERL Program

; SEQ ID NO 693

; LENGTH: 7983

; TYPE: DNA

Jones, Anissa L.

Yu, Jimmy Y.

Greenawalt, Lila B.

Panzer, Scott R.

Roseberry, Ann M.

Wright, Rachel J.

Chen, Wensheng

Liu, Tommy

Yap, Pierre E.

Stockdreher, Theresa K.

Amshey, Stefan

Fong, Willy Tuen

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 234216.15.i
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 3518
; OTHER INFORMATION: a, t, c, g, or other
US-60-184-771-693

Query Match      93.6%; Score 1635.4; DB 96; Length 7983;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1672; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

QY 60  gtccagcacataccggagcctccggaaggtgtgtgtatgtccctacacccagggaagtg 119
Db 248  gtccagcacataccggagcctccggaaggtgtgtgtatgtccctacacccagggaagtg 307

QY 120  ggaagggagctgggacccagcctggtaagcatcccccattggcccccagctcactgtgcg 179
Db 308  ggaagggagctgggacccagcctggtaagcatcccccattggcccccagctcactgtgcg 367

QY 180  tgcacaactgtcgtccatcactgaatcagacaaagtcttctatcaacaggtcccaactggga 239
Db 368  tgcacaactgtcgtccatcactgaatcagacaaagtcttctatcaacaggtcccaactggga 427

QY 240  aggcactctgggctggcctctgtgattgccaggtccagcctcccccagccttt 299
Db 428  aggcactctgggctggcctctgtgattgccaggtccagcctcccccagccttt 487

QY 300  ctttgactctctgtgaaagcagaccccagcttcccacactctctccctgcagctttgtg 359
Db 488  ctttgactctctgtgaaagcagaccccagcttcccacactctctccctgcagctttgtg 547

QY 360  tgcgtgttcccccataccagctgaagtgtcgtggtccctctgtcgtgagggagcagatgcac 419
Db 548  tgcgtgttcccccataccagctgaagtgtcgtggtccctctgtcgtgagggagcagatgcac 607

QY 420  tgaaggtatgcacacactcgtgtacacagcagctcctctgtatatacccccacccggcgga 479
Db 608  tgaaggtatgcacacactcgtgtacacagcagctcctctgtgtatatacccccacccggcgga 667

QY 480  gtggtattaggtgatcatctgtgctgggtggagatcaatggacagatctgaaatggga 539
Db 668  gtggtattaggtgatcatctgtgctgggtggagatcaatggacagatctgaaatggga 727

QY 540  ctgcaagaggtacaaactatgacaaagacattgtggacagctggcaccacacacttcgttt 599
Db 728  ctgcaagaggtacaaactatgacaaagacattgtggacagctggcaccacacacttcgttt 787

QY 600  gcccaagaaagtgttgaagctcagtcacaaatccatcaaggcagcctctccacggagaa 659
Db 788  gcccaagaaagtgttgaagctcagtcacaaatccatcaaggcagcctctccacggagaa 847

QY 660  gttccctgcaggtttctggtagagagcagctggtgtgtcgtggaagcagccaccacc 719
Db 848  gttccctgcaggtttctggtagagagcagctggtgtgtcgtggaagcagccaccacc 907

QY 720  ttggaaacattttccacgtatctactctactctactctactctactctactctactctactct 779
Db 908  ttggaaacattttccacgtatctactctactctactctactctactctactctactctactct 967

QY 780  ccgcataccactctccgcagacaaatccctgcgcagctggaagatgtgccacgtccca 839
Db 968  ccgcataccactctccgcagacaaatccctgcgcagctggaagatgtgccacgtccca 1027

QY 840  agacgactgttacaagtgttccatctcacagtcacagtcacagtcacagtcacagtcacagtcg 899
Db 1028  agacgactgttacaagtgttccatctcacagtcacagtcacagtcacagtcacagtcacagtcg 1087

QY 900  tatcatggagggtttctactgttcttctgtatc--gggcccggaaaaacgaattgtgttctgtg 958

```

```

Db 1088  tateatggagggtcttctactgttctcttgcgtggggccgcaaaacgaattgtgttctgtg 1147
QY 959  tcagcgtctgccatgtcacagatagttcagggacgagcgtgtggaaggccnctttgtc 1018
Db 1148  tcagcgtctgccatgtcacagatagttcagggacgagcgtgtggaaggccnctt-tttgtc 1206
QY 1019  accttggacatggaagactgtggtacaaacttccacagacagatgagtcaaccctcatg 1078
Db 1207  accttggacatggaagactgtggtacaaacttccacagacagatgagtcaaccctcatg 1266
QY 1079  accatagcctatgtcatgtggtccatctgcgcctctcttcactgtgcacactgtcctcatg 1138
Db 1267  accatagcctatgtcatgtggtccatctgcgcctctcttcactgtgcacactgtcctcatg 1326
QY 1139  gtgtgtcagtggtcgtcctccgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1198
Db 1327  gtgtgtcagtggtcgtcctccgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1386
QY 1199  atctccctgtgaagtgaaggagggcccatgggcagaaatagggattccctccctggaccacac 1258
Db 1387  atctccctgtgaagtgaaggagggcccatgggcagaaatagagattccctccctggaccacac 1446
QY 1259  ctccgtggttcaacttgggtcacaaagttaggagacacagatggcaccctgtgcccagagcacc 1318
Db 1447  ctccgtggttcaacttgggtcacaaagttaggagacacagatggcaccctgtgcccagagcacc 1506
QY 1319  tcaggaccctcccccacccacaaatgcctcgtcgttgcgtgagaaagaaagcctggcga 1378
Db 1507  tcaggaccctcccccacccacaaatgcctcgtcgttgcgtgagaaagaaagcctggcga 1566
QY 1379  ggtgggttccagggactgtacctgtaggagacacagaaagaaagaaagcactctgct 1438
Db 1567  ggtgggttccagggactgtacctgtaggaaacacagaaagaaagaaagcactctgct 1626
QY 1439  ggcgggaactactctgtgtcaccctcaaatcttaagtgcggaaattctgtcgttgaacttc 1498
Db 1627  ggcgggaactactctgtgtcaccctcaaatcttaagtgcggaaattctgtcgttgaacttc 1686
QY 1499  agccctgaaccttgt--caccattcctttaaattctccaccccaaaagtattctctttc 1557
Db 1687  agccctgaaccttgtcaccattcctttaaattctccaccccaaaagtattctctttc 1746
QY 1558  ttagtctcagaagtactgtggtcacacgcaggttaccttggcgtgtgtcctcgttgcgttacc 1617
Db 1747  ttagtctcagaagtactgtggtcacacgcaggttaccttggcgtgtgtcctcgttgcgttacc 1806
QY 1618  ctggcagagaagagacccaagcttgttccctcgtcgtgcccagaagtcagtaggagaggatgca 1677
Db 1807  ctggcagagaagagacccaagcttgttccctcgtcgtgcccagaagtcagtaggagaggatgca 1866
QY 1678  cagttgtctatttgccttagagacagggactgtataaacagcctaactgtgtgcaaaa 1737
Db 1867  cagttgtctatttgccttagagacagggactgtataaacagcctaactgtgtgcaaaa 1926
QY 1738  a 1738
Db 1927  a 1927

RESULT 10
US-09-009-191-1
; Sequence 1, Application US/09009191
; GENERAL INFORMATION:
; APPLICANT: POWELL, DAVID
; APPLICANT: CHAPMAN, CONRAD
; APPLICANT: MURPHY, KAY
; APPLICANT: SMITH, TRUDI
; TITLE OF INVENTION: ASP2
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE

```


|||||
Db 1970 CAGTTTCTATTGCTTTAGAGACAGGACTGTATATAACAAGCTATACATTTGGTGCAAG 2029

Qy 1738 a 1738

Db 2030 A 2030

RESULT 11

US-09-009-191-3

; Sequence 3, Application us/09009191

; GENERAL INFORMATION:

; APPLICANT: POWELL, DAVID

; APPLICANT: CHAPMAN, CONRAD

; APPLICANT: MURPHY, KAY

; APPLICANT: SMITH, TRUDI

; TITLE OF INVENTION: ASP2

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: RATNER & PRESTIA

; STREET: P.O. BOX 980

; CITY: VALLEY FORGE

; STATE: PA

; COUNTRY: USA

; ZIP: 19482

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/009,191

; FILING DATE: 20-JAN-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: UK 9701684.4

; FILING DATE: 28-JAN-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: PRESTIA, PAUL F

; REGISTRATION NUMBER: 23,031

; REFERENCE/DOCKET NUMBER: GH-70368

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-407-0700

; TELEFAX: 610-407-0701

; TELEX: 846169

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2370 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

; US-09-009-191-3

Query Match 92.2%; Score 1611.4; DB 35; Length 2370;

Best Local Similarity 98.6%; Pred. No. 0;

Matches 1657; Conservative 0; Mismatches 21; Indels 3; Gaps 3;

Qy 60 gtccagcacataccgggacccctccggaagggtgtgtatgtgacctacacccaggggcaagt 119

Db 180 GTCCAGCACATACCGGGACCTCCGGAAGGTTGTATGAGCCCTACACCCAGGGCAAGTG 239

Qy 170 ggaagggagctgggcaccagcctgttaagcatcccccatggcccaacgctcactgtgcg 179

Db 240 GGAAGGGAGCTGGGCACCCGACCTGGTAAGCATCCCCCATGGCCCCAACGTCACGTGCGC 299

Qy 180 tgcacacattgtgcacactgaatcacagaagtcttctcatcaacggctccaactggga 239

Db 300 TGCCAACATTGCTGCCATCACTGAATCAGACAAGTTCTTCATCAACGGCTCCAACTGGGA 359

Qy 240 aggcactctggggtgcctatgtgtgagattgccaggcctgcacgactccccggggccttt 299

Db 360 AGSCATCCTGGGCTGGCCTATGCTGAGATTGCCAGGCTGACGACTCCCTGGAGCCTTT 419

Qy 300 ctttgactcttgtaaaagcagaccacagttcccaacctctctccctcagctttgtgg 359

Db 420 CTTTGACTCTCTGGTAAAGCAGACCCACGTTCCCAACCTCTTCTCCCTGAGCTTTGTGG 479

Qy 360 tgcctgcttccccctcaaccagctgaagtgcctgtgcggagggagagatgatcat 419

Db 480 TGCTGGCTTCCCCTCAACCACTGCTGAAGTGTGGCTCTGTGCGGAGGAGCATGATCAT 539

Qy 420 tggagtgatcgaccactcgctgtacacagcgactctctgtgtatacaccccatccggcgagga 479

Db 540 TGGAGGTATCGACCACTCGCTGTACACAGCAGCTCTGTGTATACACCCATCCGCGGGGA 599

Qy 480 gtggtattatgaggtgatcatctgcgggtggagatcaatggacagatctgaaatgga 539

Db 600 GTGGTATTATGAGGTGATCATTTGCGGGTGGAGATCAATGGACAGGATCTGAAATGGA 659

Qy 540 ctgcaaggaggtacaaactatgacaagagcattgtggacagtggcaccaccacacctctgttt 599

Db 660 CTCGAGGAGTACAACATATGACAAGAGCATTTGGACAGTGGCACCACCACCACTTCGTTT 719

Qy 600 gcccaagaaagtgtttgaagctcagtcataatccatcaaggcgctctccacagagaa 659

Db 720 GCCCAAGAAAGTGTTTGAAGCTGAGTCAAAATCCATCAAGCGACGCTCTCCAGGGGAGAA 779

Qy 660 gtccctgacggtttctggctagagacagctggtgtctgcgaagcagcagcaccacccc 719

Db 780 GTTCCCTGATGGTTCCTGGCTAGAGACAGCTGGTGTCTGCAGACGACGACACACCCC 839

Qy 720 ttggaaacttttcccagtcactctacctaattgggtgaggttaccacacacagtcctt 779

Db 840 TTGGAACATTTTCCCAGTCATCTCACTTACCTAAATGGGTGAGTTACCAACCACTCCTT 899

Qy 780 ccgcataccatccttcgcagcaatactcgcgcagtggaagatgtgccaagcagtcacca 839

Db 900 CCGCATCACCATCTTCCGCGAGCAATACCTGCGGCCAGTGGAAAGATGTGCCACGTCCTCA 959

Qy 840 agacgactgtacaagtttgcctctcacagtcacacggtcacggtgactgttatggagctgt 899

Db 960 AGACGACTGTTTACAAGTTTGCCTATCTCACAGTCATCCACGGGCACGTGTTATGGAGCTGT 1019

Qy 900 tatcatggagggtcttctacgttcttcttgcctcgggcccgaacgaattggctttgtgt 959

Db 1020 TATCATGGAGGGCTTCTACGTTGTCTTTGATCGGCCCCGAAACGAATTGGCTTGTGTGT 1079

Qy 960 cagcgttgccatgtgcacagatgagttcaggaacggcagcggtggaagccnctttgtca 1019

Db 1080 CAGCGCTTGCCTATGTGCACGATGAGTTTCAGGCGGCGAGCGGTGGAAAGGCCCT - TTTTGTCA 1138

Qy 1020 ccttggacatgggaagactgtggtcacacaattccacagacagatgagtcacacccctatga 1079

Db 1139 CCTTGGACATGGAGACTGTGGCTACAAACATTCACAGACAGAGATGAGTCAACCTCATGA 1198

Qy 1080 ccatagctatgtoatggctgacctcgcctcttctcatcgtgcacctctgcctcatgg 1139

Db 1199 CCATAGCCTATGTATGCTGCCATCTGCGCCCTCTTTCATGCTGCCACTCTGCTCATAGG 1258

Qy 1140 tgtgtcagtgggcgtgcctccgctgcctgcgcagcagcatgatgacttttgtgatgaca 1199

Db 1259 TGTGTAGTGGCGTGCCTCCGCTGCGCTGCGCCAGACAATGGATGACTTTGCTGTGATGACA 1318

Qy 1200 tctccctgctgaagtggaggggcccatgggcagaagataggattccctt - ggaccacac 1258

Db 1319 TCTCCCTGCTGAAGTGGAGAGGCCCATGGGAGAAAGATAGAGATTCCCTGGGACCACAC 1378

Qy 1259 ctcogtggttcactttgtgtacaaagttagggagacacagatggcacctgtggccagagcacc 1318

Db 1379 CTCGCTGGTTTCACTTTGGTCAAAAGTAGGAGACAGATGSCACCTGTGGCCAGAGCACC 1438

Qy 1319 tcaggacacctccccaccaccacaaatgctctgtgtgaggaaggaagagctggcaca 1378

Db 1439 TCAGGACCCCTCCCCACCACCAAAATGCTCTGCTGCTGATGGAGGAAGGAAGGCTGGCAA 1498

Db 1475 tccgtggttcaacttttggtacaaagttaggagacacagatggccacctgtgcccagagcaact 1534
QY 1320 caggacctcccccacccacccaaatgctctgcttgatggaagaaaggctggaag 1379
Db 1535 caggacctcccccacccacccaaatgctctgcttgatggaagaaaggctggaag 1594
QY 1380 gtgggttccaggactgtaccttagagacagaaagaaagaaagacactctgtg 1439
Db 1595 gtgggttccaggactgtaccttagaaacagaaagaaagaaagacactctgtg 1654
QY 1440 gcgggaatacttggctcacctcaaatlaagtcgggaattctgcttgaaacttca 1499
Db 1655 gcgggaatacttggctcacctcaaatlaagtcgggaattctgcttgaaacttca 1714
QY 1500 gccctgaacctttgt-caccattcctttaaattctccaaacccaaagtattctctttct 1558
Db 1715 gccctgaacctttgtccaccattcctttaaattctccaaacccaaagtattctctttct 1774
QY 1559 tagttcagaagtactggaatacacagcaggttaaccttggcgtgtgtccctgtggtaccc 1618
Db 1775 tagttcagaagtactggaatacacagcaggttaaccttggcgtgtgtccctgtggtaccc 1834
QY 1619 tggcagagaagacacaaagtgttttccctgctggccaaagtcaagtaggagagatgcac 1678
Db 1835 tggcagagaagacacaaagtgttttccctgctggccaaagtcaagtaggagagatgcac 1894
QY 1679 agttgtattgtcttagagacaggagactgtataaacagcctaactgtgtgcaaaaa 1738
Db 1895 agttgtattgtcttagagacaggagactgtataaacagcctaactgtgtgcaaaaa 1954

RESULT 13
US-09-548-366-5
; Sequence 5, Application US/09548366
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; FILE REFERENCE: 28341/6280A
; CURRENT APPLICATION NUMBER: US/09/548,366
; CURRENT FILING DATE: 2000-04-12
; PRIORITY APPLICATION NUMBER: 60/155,493
; PRIORITY FILING DATE: 1999-09-23
; PRIORITY APPLICATION NUMBER: 09/404,133
; PRIORITY FILING DATE: 1999-09-23
; PRIORITY APPLICATION NUMBER: PCT/US99/20881
; PRIORITY FILING DATE: 1999-09-23
; PRIORITY APPLICATION NUMBER: 60/101,594
; PRIORITY FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-548-366-5

Query Match 85.38; Score 1489.6; DB 57; Length 1977;
Best Local Similarity 95.28; Pred. No. 0;
Matches 1599; Conservative 0; Mismatches 4; Indels 77; Gaps 3;

QY 60 gtccagcacataccgggacccctccggaagggtgtgtatgtgcccatacccccagggaagt 119
Db 351 gtccagcacataccgggacccctccggaagggtgtgtatgtgcccatacccccagggaagt 410
QY 120 ggaagggaagtgggacccgacctggtaagcatcccccatcccccaactgcaactgtgag 179

Db 411 ggaagggaagtgggacccgacctggtaagcatcccccatcccccaactgcaactgtgag 470
QY 180 tgccaacattgtgcacactcaactgaatcagaagaagttcttcatcaacgctcacaactggga 239
Db 471 tgccaacattgtgcacactcaactgaatcagaagaagttcttcatcaacgctcacaactggga 530
QY 240 aggcatacctggggctgggacctatgctgagattgccaggcctgacgactccccgggagccatt 299
Db 531 aggcatacctggggctgggacctatgctgagattgccaag----- 567
QY 300 ctttgactctctggttaaagcagacccacggttcccaacctctctccctgcagctttgtgg 359
Db 567 -----gctttgtgg 575
QY 360 tctggcttcccccctcaaccagctgaagtgtgacctgtcgagggagcatgatcat 419
Db 576 tctggcttcccccctcaaccagctgaagtgtgacctgtcgagggagcatgatcat 635
QY 420 tggaggtatcgaccactcgtgtacacagcagctctctgtgtatatacacccatccggcgga 479
Db 636 tggaggtatcgaccactcgtgtacacagcagctctctgtgtatatacacccatccggcgga 695
QY 480 gtggattattgagggtgacttctgtcggggtggagatcaatggacagggatctgaataatgga 539
Db 696 gtggattattgagggtgacttctgtcggggtggagatcaatggacagggatctgaataatgga 755
QY 540 ctgcaaggaggtacaactatgacaagagcattgtggacagtgggcaccacacactctcgttt 599
Db 756 ctgcaaggaggtacaactatgacaagagcattgtggacagtgggcaccacacactctcgttt 815
QY 600 gcccaagaaagtgtttgaagctgcagtcataatccatcaaggcagcctctccacaggaagaa 659
Db 816 gcccaagaaagtgtttgaagctgcagtcataatccatcaaggcagcctctccacaggaagaa 875
QY 660 gttccctgacgggtttctgctgaggagagacagctgggtgctggaagcagggacacacccc 719
Db 876 gttccctgaggtttctgctgaggagagacagctgggtgctggaagcagggcaccacccc 935
QY 720 ttggaaacttttccacgtcatctcactctacctaagtggtgaggttaccacacagtcctt 779
Db 936 ttggaaacttttccacgtcatctcactctacctaagtggtgaggttaccacacagtcctt 995
QY 780 ccgcatcacatcttcccgacgcaatctctgcccagtcagtcagtcagtcagtcagtcagtcag 839
Db 996 ccgcatcacatcttcccgacgcaatctctgcccagtcagtcagtcagtcagtcagtcagtcag 1055
QY 840 agacgactgttacaagtttgccatctcacagtcacagtcacagtcacagtcacagtcagtcag 899
Db 1056 agacgactgttacaagtttgccatctcacagtcacagtcacagtcacagtcacagtcagtcag 1115
QY 900 tatcatggagggtcttctacgtttgtcttctgacggggcccgaaacgaattggctttgtctgt 959
Db 1116 tatcatggagggtcttctacgtttgtcttctgacggggcccgaaacgaattggctttgtctgt 1175
QY 960 cagcgttgcctatgtgcacagatgagttcagggacggcagcgggtggaagggcccttttgtca 1019
Db 1176 cagcgttgcctatgtgcacagatgagttcagggacggcagcgggtggaagggccc-ttttgtca 1234
QY 1020 ccttgacatgggaagactgtggtacaacattccacagacagatgagtcacacccctcatga 1079
Db 1235 ccttgacatgggaagactgtggtacaacattccacagacagatgagtcacacccctcatga 1294
QY 1080 ccatagcctatgcatgggtgccatctgcgcctcttctatgctgccactctgctcctcatgg 1139
Db 1295 ccatagcctatgcatgggtgccatctgcgcctcttctatgctgccactctgctcctcatgg 1354
QY 1140 tgtgtcaagtggcctgcctccgctgctgcgcagcagcatgatgactttgtcgaagaca 1199
Db 1355 tgtgtcaagtggcctgcctccgctgctgcgcagcagcatgatgactttgtcgaagaca 1414
QY 1200 tctccctgctggaagtggagggcccatgggcagaaagatagggttccctctggaccacacc 1259
Db 1415 tctccctgctggaagtggagggcccatgggcagaaagatagggttccctctggaccacacc 1474


```
QY 1260 tccgtggttcaacttgggtcacaagttagagacacagatggcacctgtggccagagacacct 1319
D 1475 tccgtggttcaacttgggtcacaagttagagacacagatggcacctgtggccagagacacct 1534
QY 1320 caggacctcccccacacacaaatgcctctgccttgatgagaaaggaagggcggcaag 1379
D 1535 caggacctcccccacacacaaatgcctctgccttgatgagaaaggaagggcggcaag 1594
QY 1380 gtgggttccagggaacttaactgttagagacagaaaagagaagaaaagaacactctgtg 1439
D 1595 gtgggttccagggaacttaactgttagaacaagaaaagagaagaaaagaacactctgtg 1654
QY 1440 gcgggaataacttgggtcaactcaaatttaagtgcggaaattctgcttgaaacttca 1499
D 1655 gcgggaataacttgggtcaactcaaatttaagtgcggaaattctgcttgaaacttca 1714
QY 1500 gccctgaacctttgt-caccattctttaattctcaacccaaagtattcttttct 1558
D 1715 gccctgaacctttgtcacaattctttaattctcaacccaaagtattcttttct 1774
QY 1559 tagtttcagaagtactggcatcacacgcaggttaccttggcgtgtgtccctgtgtgtaacc 1618
D 1775 tagtttcagaagtactggcatcacacgcaggttaccttggcgtgtgtccctgtgtgtaacc 1834
QY 1619 tggcagagaagagacaaagtgtttccctgtcgtggccaaagtcaagtaggagagatgcac 1678
D 1835 tggcagagaagagacaaagtgtttccctgtcgtggccaaagtcaagtaggagagatgcac 1894
QY 1679 agtttgctatttcttagagacagggactgtataaaacagcctaacttggtgcacaaa 1738
D 1895 agtttgctatttcttagagacagggactgtataaaacagcctaacttggtgcacaaa 1954
```

RESULT 14

```
US-09-548-367-5
; Sequence 5, Application US/09548367
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES THEREFOR
; FILE REFERENCE: 28341/6280A
; CURRENT APPLICATION NUMBER: US/09/548,367
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-548-367-5
```

```
Query Match 85.3%; Score 1489.6; DB 57; Length 1977;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 1599; Conservative 0; Mismatches 4; Indels 77; Gaps 3;
```

```
QY 60 gtccagcacataccggacctccggaaaggtgtgtatgtgcccctacacccaggcgaagt 119
D 351 gtccagcacataccggacctccggaaaggtgtgtatgtgcccctacacccaggcgaagt 410
```

```
QY 120 ggaaggaggagctgggcccagacctgggtaagcatcccccatggcccccaacgtcaactgtgc 179
D 411 ggaaggaggagctgggcccagacctgggtaagcatcccccatggcccccaacgtcaactgtgc 470
QY 180 tgcaacaattgtcctccactcaactgaatcagacaagtctctcatcaacgctccaactggga 239
D 471 tgcaacaattgtcctccactcaactgaatcagacaagtctctcatcaacgctccaactggga 530
QY 240 aggcatacctggggctggcctatgctgagattggcaggcctgacgactccccggagccttt 299
D 531 aggcatacctggggctggcctatgctgagattggcaggcctgacgactccccggagccttt 567
QY 300 ctttgactctctgtaaaagcagacccacgcttcccaacctcttctccctgcagcctttgg 359
D 567 -----gctttgtgg 575
QY 360 tgcgtgcttccccctcaaccagctctgaagtctgctggcctctgcgaggaggagcatgatcat 419
D 576 tgcgtgcttccccctcaaccagctctgaagtctgctggcctctgcgaggaggagcatgatcat 635
QY 420 tggagggtatcgaccactcgtgttacacaggcagctctctgttatcacaccatccgcgggga 479
D 636 tggagggtatcgaccactcgtgttacacaggcagctctctgttatcacaccatccgcgggga 695
QY 480 gtggtattatgagtgatcatctgtcgggtggagataaatgagcaggatctgaaaaatgga 539
D 696 gtggtattatgagtgatcatctgtcgggtggagataaatgagcaggatctgaaaaatgga 755
QY 540 ctgcaaggagtcacaactatgacaagagcattgtgacagtgagcagcaccacacacctcttt 599
D 756 ctgcaaggagtcacaactatgacaagagcattgtgacagtgagcagcaccacacacctcttt 815
QY 600 gcccaagaaagtgtttgaagctgcagtcacaaatccatcaagcagcctcctccacggagaa 659
D 816 gcccaagaaagtgtttgaagctgcagtcacaaatccatcaagcagcctcctccacggagaa 875
QY 660 gtccctgcagcgtttctgcttaggagagcagctggtgtcgtggcaagcaggacacacccc 719
D 876 gtccctgcagcgtttctgcttaggagagcagctggtgtcgtggcaagcaggacacacccc 935
QY 720 ttgaaacatttcccagtcactctcactacctaagtgggtgaggttaccacacacctctt 779
D 936 ttgaaacatttcccagtcactctcactacctaagtgggtgaggttaccacacacctctt 995
QY 780 ccgcatcaccatccttcgcgagcaatacctgcgcagtggaagtgtggccagctccca 839
D 996 ccgcatcaccatccttcgcgagcaatacctgcgcagtggaagtgtggccagctccca 1055
QY 840 agacgactgttacaagtgttgcatactcacagtcacacgggagcactgttatggagctgt 899
D 1056 agacgactgttacaagtgttgcatactcacagtcacacgggagcactgttatggagctgt 1115
QY 900 tatcatggagggtcttcacgttcttcttgatcgggcccgaagaaatggcttctgtctgt 959
D 1116 tatcatggagggtcttcacgttcttcttgatcgggcccgaagaaatggcttctgtctgt 1175
QY 960 cagcgttgcctatgtcacgatgagttcaggaagcagcagcgtggaagccctntttgca 1019
D 1176 cagcgttgcctatgtcacgatgagttcaggaagcagcagcgtggaagggccc-ttttgca 1234
QY 1020 ccttggacatggaagactgttggttacaacattccacagacagatgagtcacacctcatga 1079
D 1235 ccttggacatggaagactgttggttacaacattccacagacagatgagtcacacctcatga 1294
QY 1080 ccatagcctatgtcagtggtcccatctgcacctcttcatctgctgccactctgcctcatgg 1139
D 1295 ccatagcctatgtcagtggtcccatctgcacctcttcatctgctgccactctgcctcatgg 1354
QY 1140 tgtgtcagtggtcgtcctcgcctgcctgcgcagcagcagcagcagcagcagcagcagcagc 1199
D 1355 tgtgtcagtggtcgtcctcgcctgcctgcgcagcagcagcagcagcagcagcagcagcag 1414
QY 1200 tctccctgctgaagtgaaggaggcccatggggcagaagatagggtattccctctggaccaccc 1259
```

|||||
Db 1415 tctccctgctgaagttagggagccctcaggcagagatagagattccctcggaccacac 1474
QY 1260 tccgtggttcaacttgggtacaaagttaggagacacagatggcacctggtgccagagacact 1319
Db 1475 tccgtggttcaacttgggtacaaagttaggagacacagatggcacctggtgccagagacact 1534
QY 1320 caggacccctccacacccacaaatgcctctgccttgatggagagaaaaggctggcaag 1379
Db 1535 caggacccctccacacccacaaatgcctctgccttgatggagagaaaaggctggcaag 1594
QY 1380 gtgggttccaggactgtacctgtaggagacagaaaagagaagaagaagcactctgctg 1439
Db 1595 gtgggttccaggactgtacctgtaggaaacagaaaagagaagaagcactctgctg 1654
QY 1440 gcgggaatactcttggtcacctcaaatttaagtcgggaaattctgctgtgaacttca 1499
Db 1655 gcgggaatactcttggtcacctcaaatttaagtcgggaaattctgctgtgaacttca 1714
QY 1500 gccctgaacctttgt-caccattcctttaattctccaaaccaaagtattctctttct 1558
Db 1715 gccctgaacctttgtccacattcctttaattctccaaaccaaagtattctctttct 1774
QY 1559 tagtttcagaagtactggcatcacacaggttaccttggcgtgtgtccctgtggtaccc 1618
Db 1775 tagtttcagaagtactggcatcacacaggtttaccttggcgtgtgtccctgtggtaccc 1834
QY 1619 tggcagagaagagacaaagtatttccctgcctggccaaagtacgtagagagatgcac 1678
Db 1835 tggcagagaagagacaaagtatttccctgcctggccaaagtacgtagagagatgcac 1894
QY 1679 agtttgctattgttttagagacagggactgtataaacaagccttaacttggcgcaaaa 1738
Db 1895 agtttgctattgttttagagacagggactgtataaacaagccttaacttggcgcaaaa 1954

RESULT 15

US-09-548-368-5
; Sequence 5, Application US/09548368
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; FILE REFERENCE: 28341/6280A
; CURRENT APPLICATION NUMBER: US/09/548,368
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-548-368-5

Query Match 85.3%; Score 1489.6; DB 57; Length 1977;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 1599; Conservative 0; Mismatches 4; Indels 77; Gaps 3;

QY 60 gtccagacacataccggacctccggaagggtgtgtatgtgccctacacccagggaagtg 119
|||||

Db 351 gtccagacacataccggacctccggaagggtgtgtatgtccctacacccaggggcaagtg 410
QY 120 gaaaggggagctgggacacagacctgtgaagcatcccccattgcccacacgtcactgtgcg 179
Db 411 gaaaggggagctgggacacagacctgtgaagcatcccccattgcccacacgtcactgtgcg 470
QY 180 tgcaacattgtctgccatcactgaatcagacaagttcttcatacaacgggtccaactggga 239
Db 471 tgcaacattgtctgccatcactgaatcagacaagttcttcatacaacgggtccaactggga 530
QY 240 aggcattcctgggctggcctatgtctgagattgccaggcctgacgactcccgagagccttt 299
Db 531 aggcattcctgggctggcctatgtctgagattggca----- 567
QY 300 ctttgactctctgtgtaagcagacccacgttcccaacctcttctccctgcagactttgttg 359
Db 567 -----gctttgttg 575
QY 360 tctgtgtctccctcaacacagctgaagtgtggtgctgtgctgtgctggaggagcagatgcat 419
Db 576 tctgtgtctccctcaacacagctgaagtgtggtgctgtgctgtgctggaggagcagatgcat 635
QY 420 tggaggtatcgacacacactgcgtgtacacagagcagctctgtgtatatacaccatccggcgga 479
Db 636 tggaggtatcgacacacactgcgtgtacacagagcagctctgtgtatatacaccatccggcgga 695
QY 480 gtgtattatgagtgatcatgtgcgggtggagatcaatggacaggtatctgaaaatgga 539
Db 696 gtgtattatgagtgatcatgtgcgggtggagatcaatggacaggtatctgaaaatgga 755
QY 540 ctgcaaggagtagcaactatgacaagagcattgtggacagtggcaccacacactctgcttt 599
Db 756 ctgcaaggagtagcaactatgacaagagcattgtggacagtggcaccacacactctgcttt 815
QY 600 gcccaagaaagtgttgaagctgcagtcacataacatccatacaggcagcctctccacagagaa 659
Db 816 gcccaagaaagtgttgaagctgcagtcacataacatccatacaggcagcctctccacagagaa 875
QY 660 gttccctgacgggttctggtgtagagagcagctggtgtgctggcagagcagggcaccacccc 719
Db 876 gttccctgaggttctggtgtagagagcagctggtgtgctggcagagcagggcaccacccc 935
QY 720 ttggaacattttcccsagtcactctcactctacctaattggtgaggttaccacacagtcctt 779
Db 936 ttggaacattttcccsagtcactctcactctacctaattggtgaggttaccacacagtcctt 995
QY 780 ccgcataccacatcttccgcagcaataacatccctgcgcagtggaagatgtgcccacgtccca 839
Db 996 ccgcataccacatcttccgcagcaataacatccctgcgcagtggaagatgtgcccacgtccca 1055
QY 840 agacgactgttacaagtgttgcacatctcacagtcacacagggcagcactgttatggagactgt 899
Db 1056 agacgactgttacaagtgttgcacatctcacagtcacacagggcagcactgttatggagactgt 1115
QY 900 tatcatggagggttctacgtgtcttcttgatcgggcccgaacacgaattggctttgctgt 959
Db 1116 tatcatggagggttctacgtgtcttcttgatcgggcccgaacacgaattggctttgctgt 1175
QY 960 cagcgttgcattgtgcacagatgattcagagcggcagcgggtggaagggccncttttgtca 1019
Db 1176 cagcgttgcattgtgcacagatgattcagagcggcagcgggtggaagggccc-ttttgtca 1234
QY 1020 ccttgacatgggaagactgtggctacaacattccacagacagatgagatcaacctcatga 1079
Db 1235 ccttgacatgggaagactgtggctacaacattccacagacagatgagatcaacctcatga 1294
QY 1080 ccatagcctatgtcatggtggtccatctgcctcttcttcattgctgacactgtgctcatgg 1139
Db 1295 ccatagcctatgtcatggtggtccatctgcctcttcttcattgctgacactgtgctcatgg 1354
QY 1140 tgtgtcagtggtgctgcctcctgcctgcctgcgcagcagcagatgatctttgtgtatgaca 1199
Db 1355 tgtgtcagtggtgctgcctcctgcctgcctgcgcagcagcagatgatctttgtgtatgaca 1414

```
Qy 1200 tctccctgctgaagtggaggcccatgggcagaaagataggattccctcctggaccacacc 1259
Db 1415 tctccctgctgaagtggaggcccatgggcagaaagataggattccctcctggaccacacc 1474
Qy 1260 tccgtggttcaacttgggtcacaagtagagacacagatggcacctgtggccagagcacct 1319
Db 1475 tccgtggttcaacttgggtcacaagtagagacacagatggcacctgtggccagagcacct 1534
Qy 1320 caggacctcccaaccacacaaatgctctgcttctgctgagagaagaaaaggctggcaag 1379
Db 1535 caggacctcccaaccacacaaatgctctgcttctgctgagagaagaaaaggctggcaag 1594
Qy 1380 gtgggttccaggaggactgtacctgtaggagacagaaagagaagaaagcactctgctg 1439
Db 1595 gtgggttccaggaggactgtacctgtaggaaacagaaagagaagaaagcactctgctg 1654
Qy 1440 gcgggaatactctgtgcaacctcaaatgctgggaaattctgctgctgaaacttca 1499
Db 1655 gcgggaatactctgtgcaacctcaaatgctgggaaattctgctgctgaaacttca 1714
Qy 1500 gccctgaacctttgt-caccttctttaaattctccaacccaagattctcttttct 1558
Db 1715 gccctgaacctttgtcaccattctttaaattctccaacccaagattctcttttct 1774
Qy 1559 tagtttcagaagtactggcatcacacgaggttaccttggcgtgtgccctgtggtaccc 1618
Db 1775 tagtttcagaagtactggcatcacacgaggttaccttggcgtgtgccctgtggtaccc 1834
Qy 1619 tggcagagaagagaccacgaactgtttccctgctggccaaagtcagtaggagagagatgcac 1678
Db 1835 tggcagagaagagaccacgaactgtttccctgctggccaaagtcagtaggagagatgcac 1894
Qy 1679 agtttgotatttggcttagagacagggactgtataaacaagcctaacattgggtgcaaaa 1738
Db 1895 agtttgotatttggcttagagacagggactgtataaacaagcctaacattgggtgcaaaa 1954
```

Search completed: May 24, 2000, 01:00:37
Job time: 14209 sec

THIS PAGE BLANK (USPTO)


```

QY 121 tgagag 126
Db 196 TGAGAG 201

RESULT 2
T93809
ID T93809 standard; DNA; 1957 BP.
AC T93809;
DE 25-FEB-1998 (first entry)
KW CRYIIIA gene; corn rootworm toxicity; CryIIIA protein;
OS increased toxicity; Coleopteran insect; Colorado potato beetle;
OS relative solvent accessibility; plant resistance; ds.
OS Synthetic.
OS Bacillus thuringiensis.
FH Location/Qualifiers
Key 13..1950
CDS /*tag= a
/product= cryIIIA protein
US5659123-A.
PD 19-AUG-1997. 295060.
PF 26-AUG-1994; US-295060.
PR (PLBZ ) PLANT GENETIC SYSTEMS NV.
PA Jansens S, Peferoen M, Van Rie J;
DR WPI: 97-424316/39.
DR P-PSDB; W34784.
PT Modified Bacillus thuringiensis CryIIIA proteins - with increased
PT toxicity against insect pests, particularly Coleopteran insects,
PT e.g. corn rootworm and Colorado potato beetle
PS Claim 11; Columns 35-38; 22pp; English.
CC The present sequence represents an artificial modified cryIIIA coding
CC region, with each "nnn" encoding an amino acid that has been identified
CC as being involved in corn rootworm toxicity. The novel CryIIIA proteins
CC encoded by the present sequence have an increased toxicity to a
CC Coleopteran insect e.g. Colorado potato beetle. The substituted amino
CC acids are located in domain II of the CryIIIA protein, and have a
CC relative solvent accessibility of at least 40%, or are a maximum
CC distance of 3 amino acids away from an amino acid having at least 40%
CC relative solvent accessibility in domain II. The modified cryIIIA genes
CC can be operably linked to regulatory sequences allowing expression and
CC then be incorporated into the genome of a plant or seed. The modified
CC CryIIIA proteins can be used to combat, or improve plant resistance
CC towards insects, particularly Coleopteran insects.
SQ Sequence 1957 BP; 467 A; 658 C; 534 G; 271 T;

Query Match 3.1%; Score 43.6; DB 1; Length 1957;
Best Local Similarity 48.8%; Pred. No. 0.058;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 699 ccaaggaagtgttccagtgtgggggagaggtacaggttcagagagcgggagaggact 758
Db 196 TGAGAG 201
QY 759 ttgctgctttcagctgagctgctgctgtatggcatgcaggcatgagctccctgcagg 818
Db 196 TGAGAG 201
QY 260 TCCCTCTCGTGTGTCCTGTGAGCTTTACACCACTTCTGACACCATCTGCCCCA 319
Db 196 TGAGAG 201
QY 819 accggtatggcgtaccatctgtttccaggggagatcctggacgttggcaccacaaggcg 878
Db 196 TGAGAG 201
QY 320 GCGAGGACCTTGGAGGCCCTTTCAGCAGCAGGTGGAGGCCCTGTATGGACCAAGATCG 379
Db 196 TGAGAG 201
QY 879 gcaagtcccccaaaagaaatacaagccacacagctgagctgagtgagacagtgagg 938
Db 196 TGAGAG 201
QY 380 CCGACTACGCAAGAACAGGCCCTTGCCGAGCTGAGGCCCTGACAGACACACCTGGAGG 439
Db 196 TGAGAG 201
QY 939 ac 940
Db 440 AC 441

Query Match 3.0%; Score 42.4; DB 1; Length 756;
Best Local Similarity 42.9%; Pred. No. 0.082;
Matches 211; Conservative 0; Mismatches 281; Indels 0; Gaps 0;

QY 170 cgctgatactgagccctctgctgtggggccagcccccacagagagcactggccctggtc 229
Db 196 TGAGAG 201
QY 230 ttccgttcggcgtatgcggctcttttcagctggttcccccggagagctgcccacgcat 289
Db 196 TGAGAG 201
QY 290 GTCCACCGGGTGTCTCGGGGACCTGCAGGCCCGCCGACAGTGGCGCTGGACCGGCTG 349
Db 196 TGAGAG 201
QY 290 gccacactgcgttttacacgcccgcctgtggcccccgcctgccttatgtttctggac 349
Db 196 TGAGAG 201
QY 350 CCGGTGTCTCCGGGACCTGCAGGCCCGCCGACAGTGGCGCTGGACCGGCTGGTCCAC 409
Db 196 TGAGAG 201
QY 350 atccgcgggttcggcggcgtgggaccttgggggaaagtggcagccggcgccgcccctgt 409
Db 196 TGAGAG 201
QY 410 CTCCGGGACCTGCAGGCCCGCCGACAGTGGCGCTGGACCGGCTGGTCCACCGGCTG 469
Db 196 TGAGAG 201
QY 410 gtcttcagagatgccagcagcttcaggaggaatgtgtacgaaacctagcggataaggcc 469
Db 196 TGAGAG 201
QY 470 GACCTGACAGGCCCGCCGAGGTGCGCTGGACCGGCTGGTCCACCGGCTGGTCCCGGAC 529
Db 196 TGAGAG 201
QY 470 tttagccggcccatctgcgagggccctcctgcagaccaggggttcttcaatggcattgg 529
Db 196 TGAGAG 201
QY 530 CAGGCCCGCCGAGGTGCGCTGGACCGGCTGGTCCACCGGCTGGTCCCGGACCTGCGAG 589
Db 196 TGAGAG 201
QY 530 tatctgagggcagagatctctaccggctgaagatcccccccttttgaagggcccgctcg 589
Db 196 TGAGAG 201
QY 590 CGCCAGGTGCGCTGGACCGGCTGGTCCACCGGCTGGTCCCGGACCTGCGAGGCCCGCAG 649
Db 196 TGAGAG 201

RESULT 3
T16766
ID T16766 standard; DNA; 756 BP.
AC T16766;
DE 24-FEB-1997 (first entry)
KW Collagen-like polymer sequence D gene 2 encoding unit C2A24C2.
KW collagen; repetitive triad motif; recombinant production;
KW photographic; medical; structural; fibre; ds.
OS Synthetic.
PN US5496712-A.
PD 05-MAR-1996. 609716.
PR 06-NOV-1990; US-609716.
PR 12-NOV-1991; US-791960.
PR 05-NOV-1992; US-972032.
PA (PROT-) PROTEIN POLYMER.
PI Cappello J, Ferrari FA;
DR WPI: 96-150728/15.
DR P-PSDB; R93248.
DR Collagen-like polymers comprising repetitive triads - produced in
PT unicellular organisms with improved characteristics, useful in, e.g.
PT photographic and medical fibres
PT Example 2; Column 17-18; 43pp; English.
PS The invention concerns collagen-like polymers having repetitive triads
CC with reduced proline content, and where glycine is the initial amino
CC acid and the subsequent amino acids are varied. The choice of triads
CC utilised in a recombinant collagen-like polymer are chosen in order to
CC affect properties such as helix stability, hydration, solubility, gel
CC point, biodegradation and immunogenicity. Triads of particular interest
CC include GAP, GPP, GAS, GPG, GAO, GSP, GLO, GPR, GPK, GAK, GAR,
CC GER, GDR, GEP, GDA, GAH and GEA. The collagen-like polymers may impart
CC new characteristics, finding wide use in photographic, medical,
CC structural and fibre applications, and are capable of being produced in
CC unicellular microorganisms at high mol. wts. and in high efficiency.
CC Plasmid DNA from several colonies were analysed for inserts contg.
CC multiple A6 DNA fragments by digestion with FokI. The present sequence
CC was identified to contain the sequenced gene 1 monomer sequence C2A24C2.
CC The sequenced collagen-like polymers are used as immunogens for the
CC prepn. of antibodies.
SQ Sequence 756 BP; 82 A; 290 C; 292 G; 92 T;

```



```

FT misc_difference 1..30
FT /tag= b
FT /note= "Substitutes native Taq N-terminus coding
FT sequence with codons preferred by E.coli"
PD EP-482714-A.
PN 29-APR-1992.
PF 22-OCT-1991; 202731.
PR 26-OCT-1990; US-602848.
PA (EAST ) EASTMAN KODAK CO.
PI Sullivan MA;
DR P-PSDB; R22602.
PT Mutant thermus aquaticus DNA polymerase genes - used for
PT producing thermostable polymerase in increased amts. in E.coli
PT cells
PS Claim 1; Page 12 and Page 16; 25pp; English.
CC This sequence is the same as the native Taq Pol gene except for
CC changes in the first 30 nucleotides. The changes are designed to
CC increase expression of Taq Pol in an E.coli host. The substitution
CC mutant is encoded by plasmid ptaq3. It was found that ptaq4
CC expressed Taq Pol activity up to 10 times that of ptaq1 (i.e. the
CC native Taq Pol coding sequence). See also Q24223-9.
SQ Sequence 2499 BP; 428 A; 849 C; 849 G; 373 T;

Query Match 3.0%; Score 42.4; DB 1; Length 2499;
Best Local Similarity 48.0%; Pred. No. 0.13;
Matches 121; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 573 ttgagaagcccgctgctgctgagccctgcagcagcacagcgcgagcccgagctga 632
DB 620 TTCTGGAGGAGTGGGGAGCGCTGGAAGCCCTCTCAAGAACCTGGACCGGCTGAAGCCCG 679
QY 633 ccttgagccagaagaataagcacaagctgcagaatccagacctgctggagctatgtcaact 692
DB 680 CCATCGCGGAGAGATCTCGGCCACATGGACGATCTGAAGCTCTCTCTGGACCTGGCCCA 739
QY 693 caqtgcccaaggaagtgtccagtgtggggcagaggtcagcagcagcgggagcgggag 752
DB 740 AGTGCGCGACCGACCTGCCCTTGGAGGTGACCTTCGCCAAAGCGGGAGCCGACCGGG 799
QY 753 aggaacttgccttcgagcctgctgcgtgcgtatggtcgcagcagcagcagcagctccc 812
DB 800 AGGGGCTTAGGGCCTTCTTGAGAGGCTTGAGTTGGCAGCCCTCTCCACGAGTTGCGCC 859
QY 813 tgcaggaccggc 824
DB 860 TTCTGGAAAGCC 871

RESULT 7
Q24225
ID Q24225 standard; DNA; 2499 BP.
AC Q24225;
DT 30-OCT-1992 (first entry)
DE Modified Taq polymerase coding sequence ptaq5.
KW bacterial codon bias; substitution mutation; PCR; E.coli;
KW polymerase chain reaction; Taq DNA polymerase; ss.
OS Thermus aquaticus.
FH Key Location/Qualifiers
FT cds 1..2499
FT /tag= a
FT /product= Taq_Pol
FT misc_difference 1..57
FT /tag= b
FT /note= "Substitutes native Taq N-terminus coding
FT sequence with codons preferred by E.coli"
PD EP-482714-A.
PN 29-APR-1992.
PF 22-OCT-1991; 202731.
PR 26-OCT-1990; US-602848.
PA (EAST ) EASTMAN KODAK CO.
PI Sullivan MA;

```

```

DR WPI; 92-142808/18.
DR P-PSDB; R22603.
PT Mutant thermus aquaticus DNA polymerase genes - used for
PT producing thermostable polymerase in increased amts. in E.coli
PT cells
PS Claim 1; Page 12 and Page 17; 25pp; English.
CC This sequence is the same as the native Taq Pol gene except for
CC changes in the nucleotides encoding the N-terminus. The changes are
CC designed to increase expression of Taq Pol in an E.coli host. The
CC substitution mutant is encoded by plasmid ptaq5. It was found that
CC ptaq5 expressed Taq Pol activity up to 10-50 times that of ptaq1
CC (i.e. the native Taq Pol coding sequence). See also Q24223-9.
SQ Sequence 2499'BP; 437 A; 846 C; 844 G; 372 T;

Query Match 3.0%; Score 42.4; DB 1; Length 2499;
Best Local Similarity 48.0%; Pred. No. 0.13;
Matches 121; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 573 ttgagaagcccgctgctgctgagccctgcagcagcacagcgcgagcccgagctga 632
DB 620 TTCTGGAGGAGTGGGGAGCGCTTGAAGCCCTCTCAAGAACCTGGACCGGCTGAAGCCCG 679
QY 633 ccttgagccagaagaataagcacaagctgcagaatccagacctgctggagctatgtcaact 692
DB 680 CCATCGCGGAGAGATCTCGGCCACATGGACGATCTGAAGCTCTCTCTGGACCTGGCCCA 739
QY 693 caqtgcccaaggaagtgtccagtgtggggcagaggtcagcagcagcgggagcgggag 752
DB 740 AGTGCGCGACCGACCTGCCCTTGGAGGTGACCTTCGCCAAAGCGGGAGCCGACCGGG 799
QY 753 aggaacttgccttcgagcctgctgcgtgcgtatggtcgcagcagcagcagcagctccc 812
DB 800 AGGGGCTTAGGGCCTTCTTGAGAGGCTTGAGTTGGCAGCCCTCTCCACGAGTTGCGCC 859
QY 813 tgcaggaccggc 824
DB 860 TTCTGGAAAGCC 871

RESULT 8
Q24226
ID Q24226 standard; DNA; 2523 BP.
AC Q24226;
DT 30-OCT-1992 (first entry)
DE Modified Taq polymerase coding sequence ptaq6.
KW bacterial codon bias; substitution mutation; PCR; E.coli;
KW polymerase chain reaction; Taq DNA polymerase; ss.
OS Thermus aquaticus.
FH Key Location/Qualifiers
FT misc_difference 4..27
FT /tag= a
FT /note= "Inserted between codons 1 and 2 of
FT native Taq polymerase"
FT
FT EP-482714-A.
PN 29-APR-1992.
PF 22-OCT-1991; 202731.
PR 26-OCT-1990; US-602848.
PA (EAST ) EASTMAN KODAK CO.
PI Sullivan MA;
DR WPI; 92-142808/18.
DR P-PSDB; R22604.
PT Mutant thermus aquaticus DNA polymerase genes - used for
PT producing thermostable polymerase in increased amts. in E.coli
PT cells
PS Claim 1; Page 12 and Page 17; 25pp; English.
CC This sequence is the same as the native Taq Pol gene except for
CC changes in the nucleotides encoding the N-terminus. The changes are
CC designed to increase expression of Taq Pol in an E.coli host. The
CC insertion mutant is encoded by plasmid ptaq6. It was found that
CC ptaq6 expressed Taq Pol activity at least 10 times that of ptaq1
CC (i.e. the native Taq Pol coding sequence). See also Q24223-9.
SQ Sequence 2523 BP; 439 A; 853 C; 857 G; 374 T;

```


Query Match	3.08;	Score 42.4;	DB 1;	Length 2523;
Best Local Similarity	48.08;	Pred. No. 0.13;		
Matches 121;	Conservative	0;	Mismatches 131;	Indels 0;
Gaps	0;			

QY	573	ttgagaagcccgctcggtctgagcctgagcgcagcagccgagccgagctga	632
DB	644	TTCTGGAGGAGTGGGGAGCCTGGAAAGCCCTCTCAAGAACCTGGACCGGTGAAGCCG	703
QY	633	ccctgagccagaagataagaccgaagctgcagaatccagacctgctgagctatgtctact	692
DB	704	CCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTTCTCTGGACCTGGCC	763
QY	693	cagtgcccaaggaagtgcgtccagttgggggggcaagagctacggtcagagacggggaagg	752
DB	764	AGGTGCGCACCGACTGCCCTTGAGGTGGACTTCGCCAAAGCGGGAGCCGACCGGG	823
QY	753	aggacttgcctgttcctgagcctgcgtgcgtgctatggcatgccagggcatgagctccc	812
DB	824	AGGGCCTAGGCGCCTTCTGTGAGAGGCTTGAGTTTGGCAGGCTCTCTCCACGAGTTCGSCC	883
QY	813	tgcaggaccgggc	824
DB	884	TTCTGGAAAGCC	895

RESULT	9
V65798	
ID	V65798 standard; DNA; 969 BP.
AC	V65798;
DT	02-FEB-1999 (first entry)
DE	Thermus aquaticus nuclease clone 3E (Cleavase BX) DNA.
KW	Nucleic acid detection; multiple sequential invasive cleavage;
KW	DNA polymerase; nuclease; Cleavase BX; ds.
OS	Thermus aquaticus strain YT-1.
CS	Synthetic.
PN	W09842873-AL.
PD	01-OCT-1998.
PR	24-MAR-1997; U05809.
PR	24-MAR-1997; US-823516.
PA	(THIR-) THIRD WAVE TECHNOLOGIES INC.
PI	BROW MAD, Hall JG, Kwiatkowski RW, Lyamichev VI,
PI	Mast AL, Vavra SH;
DR	WPI; 98-557036/47.
PT	Detecting target nucleic acid by sequence-specific cleavage of
PT	complex with two specific oligonucleotides - used to detect
PT	cytomegalovirus DNA
PT	
PS	Example 2; Page 282-283; 524pp; English.
CC	This nucleotide sequence (clone 3E) codes for a thermostable
CC	nuclease (Cleavase BX) derived from the DNA polymerase (DNAP) of
CC	Thermus aquaticus (Tag). In comparison to the wild-type Tag DNAP
CC	sequence (see V65779), it contains an out-of-frame deletion of 1540
CC	nucleotides. The invention relates to means for the detection and
CC	characterisation of nucleic acid sequences, and variations in
CC	nucleic acid sequences. It also relates to methods for forming a
CC	nucleic acid cleavage structure on a target sequence and cleaving
CC	this structure in a site-specific manner, preferably using a
CC	thermostable structure-specific nuclease such as a modified Tag
CC	DNAP that has reduced synthetic activity (see V65783-86). Cleavage
CC	of the cleavage structure by the nuclease indicates the presence of
CC	specific nucleic acid sequences or specific variants. The invention
CC	further relates to methods for the separation of nucleic acid
CC	molecules based on charge, methods for the detection of non-target
CC	cleavage products via the formation of a complete and activated
CC	protein binding region, and methods for the detection of nucleic
CC	acid from various viruses (e.g. human cytomegalovirus) in a sample.
CC	The method amplifies the detection molecule rather than the target
CC	itself, is less subject to contamination than exponential
CC	amplification processes, and allows many targets to be analysed in
CC	a single reaction.
SQ	Sequence 969 BP; 177 A; 316 C; 327 G; 149 T;

Query Match 3.0%; Score 41.8; DB 1; Length 969;
Best Local Similarity 46.2%; Pred. No. 0.13;
Matches 139; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 524 ggaactatctgcgggagagatcgttacccgctgaagatccccccctttgagaaggcc 583
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
610 GACAACTTCCCGGGTCAGGGCATCGGGGAGAAGACGGCGAGAGACTTCTGGAGAG 669
QY 584 cgcctcgcttcctggaggccctgcagcacacagcaggccgagccggagctgacctgagccag 643
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
670 TGGGGAGCCTTGAAGCCCTCTCAAGAACCTGCAGCGGTGAAGCCGCCATCCGGGAG 729
QY 644 aagataaggaccaagctgcagaatccagaccctgctggagcttatctcactcagtcccacaag 703
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
730 AAGATCTCGGCCACATGGAGATCTGAAGCTCTCTGGGACCCTGGCCCAAGTGCGCACCC 789
QY 704 gaagtggtccagttggyggggcagaggtctacgggtcagagagcgggagagagacttctgct 763
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
790 GACCTGCCCTGGAGTTGACTTCGCCAAGGGCGGAGCCCGACCGGGAGAGGCTTAGG 849
QY 764 gccttcagacctgctgctgctatggcatgccagcgcagctcctccgcgagccagg 823
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
850 GCCTTTCTGTGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGC 909
QY 824 C 824
Db |
910 C 910

RESULT 10
Q43032
ID Q43032 standard; DNA; 756 BP.
AC Q43032;
DT 07-SEP-1993 (first entry)
DE Collagen-like polymer DCP3-C2(AB12)C2 coding sequence.
KW Recombinant; collagen-like polymers; CLP; tripeptide; helix; membrane;
KW fibre; film; coating; triad sequence; collagen; mammalian; moulding;
KW hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.
OS Synthetic.
PN W09310154-A.
PD 27-MAY-1993.
PF 04-NOV-1992; U09485.
PR 12-NOV-1991; US-791960.
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI Cappello J, Ferrari FA;
DR WPI; 93-182496/22.
DR P-PSDB; R37738.
PT High mol. wt. collagen-like protein polymers - capable of being produced in unicellular microorganisms
PS Disclosure; Page 39; 82pp; English.
CC The sequences given in Q43030-34 encode examples of recombinantly produced DCP collagen-like polymers (CLPs) which consist of repeated tripeptide sequences selected from a wide range of GXY sequences, where x and y can be any amino acid. These sequences can be cloned into plasmids and used to transform E. coli to produce the DCP proteins. DCP peptides comprise repeated units of: A = GAPGPAGPP, B = GSRGDPPP and/or C = GAHPGAPGX. These polymers may be used to raise anti-DCP antibodies in rabbits. These polymers have molecular weights of >30 kD and are able to form helices due to interchain linkages. These polymers pref. contain a proportion of tripeptide triad sequences found in natural collagens, pref. mammalian collagens. The CLPs impart unique characteristics to materials such as fibres, membranes, films, coatings, hydrogels, colloid suspensions and moulded articles.

SQ Sequence 756 BP; 84 A; 288 C; 291 G; 93 T;

Query Match 3.0%; Score 41.4; DB 1; Length 756;
Best Local Similarity 42.5%; Pred. No. 0.15; Indels 0; Gaps 0;
Matches 219; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

QY 170 cgcctgatctagccctctcgtctgggcccagccccaacaggagccactgacctgac 229

Query Match	2.9%;	Score 40.8;	DB 1;	Length 948;
Best Local Similarity	47.6%;	Pred. No. 0.23;		
Matches 120;	Conservative	0;	Mismatches 132;	Indels 0; Gaps

QY	573	ttgagaagcccgctcggtctctgagggccctgcagcagcacagccgagccgagctga	632
DB	659	TTCTGGAGGAGTGGGGAGAGCTTGAAGCCCTCCCTCAAGAACTTGACCGCTGAAGCCG	718
QY	633	ccctgagccagaagataaagaccacgaagctgcagaatccagacctgctggagctatgtca	692
DB	719	CCATCCGGGAGAAAGATCTCTGGCCACATGGACGATCTGAAGCTCTCTCTGGACCTGGCCA	778
QY	693	cagtgcccaaggaagtgtccagttggggggcagaggtacaggttcagagagcgaggag	752
DB	779	AGGTGCGCACCGACTGCCCTTGAGGTGGACTTCGCCAAAAGCGGGAGCCGACCGGG	838
QY	753	aggacttgccttcagccttcgagcctggctgcgctgctatggcatccagggcatgagctcc	812
DB	839	AGAGCCTTAGGCGCTTTCTGGAGAGCGCTTGAGTTTGGCAGGCTCTCTCCACGAGTTCGGCC	898
QY	813	tgcagggaccggc	824
DB	899	TTCTGGAAGGCC	910

RESULT 13

T70349 standard; DNA; 948 BP.

ID	T70349	
AC	T70349:	
DT	03-APR-1998 (first entry)	
DE	Synthesis deficient taq DNA polymerase gene.	
KW	DNA polymerase; taq polymerase gene; DNAP; 5' nuclease activity;	
KW	Cleavage BN; DNA cleavage; reduced synthetic activity; ss.	
OS	Synthetic.	
OS	Thermus aquaticus.	
PN	US5614402-A.	
PD	25-MAR-1997.	
PF	06-JUN-1994; 254359.	
PR	07-DEC-1992; US-986330.	
PR	04-JUN-1993; US-073384.	
PA	(THIR-) THIRD WAVE TECHNOLOGIES INC.	
PI	Brow MAD, Dahlberg JE, Lyamichev VI;	
DR	WPI; 97-201481/18.	
PT	Thermostable 5' nuclease derived from thermostable polymerase - has	
PT	reduced synthetic activity useful in nucleic acid detection assays	
PS	Claim 9; Columns 91-92; 93pp; English.	
CC	The present sequence encodes a Taq DNA polymerase enzyme that has	
CC	normal 5' nuclease activity, but reduced synthetic ability. This	
CC	sequence encodes a polymerase that is a variant of that encoded by	
CC	T70346, which itself is derived from the mutraq construct described	
CC	in T70343. The mutant polymerase encoded by the present sequence is	
CC	referred to as Cleavage BN.	
SQ	Sequence 948 BP; 179 A; 321 C; 308 G; 140 T;	

Query Match	2.9%;	Score 40.8;	DB 1;	Length 948;
Best Local Similarity	47.6%;	Pred. No. 0.23;		
Matches 120;	Conservative	0;	Mismatches 132;	Indels 0; Gaps

QY	573	ttgagaagcccgctcggtctctgagggccctgcagcagcacagccgagccgagctga	632
DB	659	TTCTGGAGGAGTGGGGAGAGCTTGAAGCCCTCCCTCAAGAACTTGACCGCTGAAGCCG	718
QY	633	ccctgagccagaagataaagaccacgaagctgcagaatccagacctgctggagctatgtca	692
DB	719	CCATCCGGGAGAAAGATCTCTGGCCACATGGACGATCTGAAGCTCTCTCTGGACCTGGCCA	778
QY	693	cagtgcccaaggaagtgtccagttggggggcagaggtacaggttcagagagcgaggag	752
DB	779	AGGTGCGCACCGACTGCCCTTGAGGTGGACTTCGCCAAAAGCGGGAGCCGACCGGG	838
QY	753	aggacttgccttcagccttcgagcctggctgcgctgctatggcatccagggcatgagctcc	812


```
Query Match          3.3%; Score 46.8; DB 1; Length 7218;
Best Local Similarity 5.1%; Pred. No. 0.0057;
Matches 21; Conservative 215; Mismatches 172; Indels 0; Gaps 0;

QY 715 gttgggggagaggtacgtgacagagcggtgagggagagacttgcgtcttcgagc 774
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1438 GTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379

QY 775 ctggctgcgtctatgcatgcagagcagatgagctccctgcagagcggcatggcgctac 834
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1378 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1319

QY 835 catctgttccaggggagatcctgacgttgccacccaaaggcgcaagtcgccgcaaaa 894
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1318 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1259

QY 895 gaaatccaaagccacacagctgagctgagggagagcagtgaggagcgttgcctccgag 954
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1258 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1199

QY 955 caaggcccttcaggagacacagagggcaagagagagaccttcctaaaggagactcaaccca 1014
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1139

QY 1015 gcgcctgaggggaccagcctccagcagagccacagagctccacagtgcccaagagg 1074
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1138 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1079

QY 1075 gaggaggaaggggcgacagcgagcgtctgccaactgagagcccccggaa 1122
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1078 RRRRRRRRRRRATCGCAAGCTCCTCGACCTGCAGCAGCAAGCTCGGAA 1031

RESULT 2
US-08-295-060-3
; Sequence 3, Application US/08295060
; Patent No. 5659123
; GENERAL INFORMATION:
; APPLICANT: VAN RIE, Jeroen
; APPLICANT: JANSSENS, Stefan
; APPLICANT: PEREROEN, Marix
; TITLE OF INVENTION: NEW DIABROTICA TOXINS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,060
; FILING DATE: 26-AUG-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1957 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..1947
US-08-295-060-3

Query Match          3.1%; Score 43.6; DB 1; Length 1957;
Best Local Similarity 4.8%; Pred. No. 0.023;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 699 ccaaggaagtgtccagttgggggacagaggtacagagcggtgagagcgaggaggaact 758
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 200 CCAAGGACGTGATCCAGAAGGGCATCAGCGTGGTGGCGACCTGCTGGCGGTGGTGGGT 259

QY 759 ttgctgcctttcagcctggtgcgtctgtatggcatgccaggeatgagctccctgcagg 818
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 260 TCCCTCTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 319

QY 819 acgggcatggcggtaccatctgtttccaggggggtccctgagccgttggcaccacaaagg 878
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 320 GCGAGACCCCTTGAAGGCTTCATGGACAGGTGGAGGCCCTGATGGACACAGAGATCG 379

QY 879 gcaagtcgccgcaaaaagaaatccaaaggccacagctgagctcctgagagcagagtggagg 938
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 380 CGGACTACGCCAAGAAAGAGGCCCTGGCCGAGCTGCAGGGCCTGCAGAAACAACGTGGAG 439

QY 939 ac 940
    ||
Db 440 AC 441

RESULT 3
US-08-642-255-50
; Sequence 50, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic"
US-08-642-255-50

Query Match 3.0%; Score 41.4; DB 2; Length 756;
Best Local Similarity 42.5%; Pred. No. 0.057;
Matches 219; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

```
QY 170 cgcctgataagccctctgcttggggccagcccaaacaggagccactggccctggtc 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 cgcctgctaccagggtgctccggagactcgacggccgccagctgcccctggaccggctg 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 230 ttccgttcggcatgtccgctcttttcagctggtgccccgagagagctgccaacgcat 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 gtcacacgggtgtccggagactcgacggccgccaggtggtggaccggctgtccac 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 290 gccacactgcttttacagggcccccgcctggccccggctcgccctatgttttctggac 349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 cggctgtccgggacctgcagggccggcaggtgctggaccggctgttccacccgggtg 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 350 atccgcccgttggccgtgggaaccttgggggaaagtggcagccgggcccggccctgt 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 CTCGGGACCTGCAGGCGCCGCGAGTGGCGCTGGACCGGCTGTCCACCGGGTCTCCGG 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 410 gtcttcagagattaccagcagtticaggagagaatgtgtctacaaaacctagcggataagccc 469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 GACCTGAGGCGCCGACAGGTGGCTGACCGGCTGTCTCCACCGGCTGTCTCCGGGACCTG 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 470 ttgacgggcccattctcgagggccctctggaccagagaggtttctcaatggcattggcaac 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 CAGSCCGCCAGGTGGCGCTGGACCGGCTGTCTCCACCGGCTGTCTCCGGGACCTGCAGGCC 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 530 tatctggggcagagatcctgtacccggtgaagatcccccccttttggagagcccgctcg 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 CGCAGGTGGCGCTGGACCGGCTGTCTCCACCGGCTGTCTCCGGGACCTGCAGGCCGCGAC 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 590 gtcttgagggccctgagcagcagccagccagccgagctgacctgagccagagaata 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 GTGGCCTGAGACCGGCTGTCCACCGGCTGTCTCCGGGACCTGCAGGCCCGCCAGGTGGCG 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 650 aggaacagctgcagaatccagacactgctgagct 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 CTGACCGGCTGGTCCACCGGCTGTCTCCGGGACCT 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 4

US-08-254-359A-31
Sequence 31, Application US/08254359A
Patent No. 5614402
GENERAL INFORMATION:
APPLICANT: DAHLBERG, JAMES E.
APPLICANT: LYAMICHEV, VICTOR I.
APPLICANT: BROW, MARY ANN D.
TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
TITLE OF INVENTION: DNA POLYMERASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,359A
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 06-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 948 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-254-359A-31

Query Match 2.9%; Score 40.8; DB 1; Length 948;
Best Local Similarity 47.6%; Pred. No. 0.09;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

```
QY 573 ttgagaagggccctcgctcctgagggccctgcagcagcacaggccgagcccgagctga 632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 TTCTGGAGGAGTGGGGAGCGCTTCTCAAGAACCTGGACCGGCTGAAGCCCG 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 633 ccttgagccagaataaagaccaaagctgcagaatccagacactgtagagctatgtact 692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 CCATCCGGGAGAGATCTCTGGCCACATGGACGATCTGAAGCTCTCTCTGGGACCTGGGCCA 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 693 cagtgcccaaggagtggtccagttggggggcagaggtacaggggtcagagagcgggagg 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 AGGTGCGCACCGACCTGCCCTTGGAGTGGACTTCGCCAAAAGGGGAGCGCCGACCGGG 838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 753 aggaacttgcctcttcagcctgagcctggtcgctgctatggaatgccagagcagctccc 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 839 AGAGGCTTAGGGCTTCTCTGGAGAGGCTTGAGTTTGGCAGGCTCTCTCCACGAGTTCCGCC 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 813 tgcagagaccgac 824
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 899 TTCTGGAAAGCC 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 5

US-08-471-066B-31
Sequence 31, Application US/08471066B
Patent No. 5837450
GENERAL INFORMATION:
APPLICANT: DAHLBERG, JAMES E.
APPLICANT: LYAMICHEV, VICTOR I.
APPLICANT: BROW, MARY ANN D.
TITLE OF INVENTION: 5' Nucleases Derived From Thermostable
TITLE OF INVENTION: DNA Polymerase
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,066B

;/ FILING DATE: 06-JUN-1995
;/ CLASSIFICATION: 435
;/ PRIOR APPLICATION DATA: US 08/254,359
;/ FILING DATE: 06-JUN-1994
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/073,384
;/ FILING DATE: 04-JUN-1993
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 07/986,330
;/ FILING DATE: 07-DEC-1992
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Ingolia, Diane E.
;/ REGISTRATION NUMBER: 40,027
;/ REFERENCE/DOCKET NUMBER: FORS-01800
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (415) 705-8410
;/ TELEFAX: (415) 397-8338
;/ INFORMATION FOR SEQ ID NO: 31:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 948 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ US-08-471-066B-31

Query Match 2.9%; Score 40.8; DB 3; Length 948;
Best Local Similarity 47.6%; Pred. No. 0.09;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 573 ttgagaagccctgctgctgagccctgcagcagcagccgagcccgagctga 632
Db 659 TTTGGAGGAGTGGGGAGCTTGAAGCCCTCTCAGAACCTGGACCGCTGAAGCCCG 718

QY 633 ccctgagccagaagataagaccgaagctgcagaaatccagaccctgctgagctatgtca 692
Db 719 CCATCCGGAGAGAGATCTGCGCCACATGACGATCTGAAGCTCTCTCTGGACCTGGCCA 778

QY 693 cagtgcacaggaagtggctcagttgggggagagctacgggtcagagagcgggaggg 752
Db 779 AGGTGCGACCGACCTGCCCTGGAGTGAGCTTCGCCAAAGGCGGGAGCCGCGGG 838

QY 753 agactttgcttttcagctgctgctgctgctatgctgcatgccagagcatagctccc 812
Db 839 AGAGGCTTAGGGCTTTCTGGAGAGGCTTGAGTTGGCAGCCCTCTCCACAGATTGGCC 898

QY 813 tgcaggaccggc 824
Db 899 TTTGGAAAGCC 910

RESULT 6
US-08-484-956-31
; Sequence 31, Application US/08484956
; Patent No. 5843654
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; APPLICANT: OLDENBURG, MARY C.
; APPLICANT: HEISLER, LAURA
; TITLE OF INVENTION: DETECTION OF p53 MUTATIONS
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:

;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/484,956
;/ FILING DATE:
;/ CLASSIFICATION: 435
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/402,601
;/ FILING DATE: 09-MAR-1995
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/337,164
;/ FILING DATE: 09-NOV-1994
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/254,359
;/ FILING DATE: 06-JUN-1994
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/073,384
;/ FILING DATE: 04-JUN-1993
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 07/986,330
;/ FILING DATE: 07-DEC-1992
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: CARROLL J, PETER G.
;/ REGISTRATION NUMBER: 32,837
;/ REFERENCE/DOCKET NUMBER: FORS-01801
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (415) 705-8410
;/ TELEFAX: (415) 397-8338
;/ INFORMATION FOR SEQ ID NO: 31:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 948 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ US-08-484-956-31

Query Match 2.9%; Score 40.8; DB 3; Length 948;
Best Local Similarity 47.6%; Pred. No. 0.09;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 573 ttgagaagccctgctgctgagccctgcagcagcagccgagcccgagctga 632
Db 659 TTTGGAGGAGTGGGGAGCTTGAAGCCCTCTCAGAACCTGGACCGCTGAAGCCCG 718

QY 633 ccctgagccagaagataagaccgaagctgcagaaatccagaccctgctgagctatgtca 692
Db 719 CCATCCGGAGAGAGATCTGCGCCACATGACGATCTGAAGCTCTCTCTGGACCTGGCCA 778

QY 693 cagtgcacaggaagtggctcagttgggggagagctacgggtcagagagcgggaggg 752
Db 779 AGGTGCGACCGACCTGCCCTGGAGTGAGCTTCGCCAAAGGCGGGAGCCGCGGG 838

QY 753 agactttgcttttcagctgctgctgctgctatgctgcatgccagagcatagctccc 812
Db 839 AGAGGCTTAGGGCTTTCTGGAGAGGCTTGAGTTGGCAGCCCTCTCCACAGATTGGCC 898

QY 813 tgcaggaccggc 824
Db 899 TTTGGAAAGCC 910

RESULT 7
US-08-757-653-31
; Sequence 31, Application US/08757653
; Patent No. 5843669
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: LYAMICHEV, NATASHA

; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,653
; FILING DATE: 23-JAN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02565
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 948 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-653-31

Query Match 2.9%; Score 40.8; DB 3; Length 948;
Best Local Similarity 47.6%; Pred. No. 0.09;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy	573	ttgagaagcccgctcgtgctgagccctgagcgacacagcgccgagcccgagctga	632
Db	559	TTCTGGAGAGTGGGGAGCGCTGGGAAGCCCTCCTCAAGAACCTGGACCGGCTGAAGCCG	718
Qy	633	ccctgagccagaagataagaccagctgcagaaatccagacacctgctgagctatgtca	692
Db	719	CCATCCGGGAGAAATCTTGGCCACATGACATGACATCTCCTGGACCTCTCTGGACCTGGCC	778
Qy	693	cagtgcccaaggaagtgtccagttgggggagaggtacggtcagagagcgggagg	752
Db	779	AGGTGCGCACCGACCTGCCCTGGAGTGGACTTGGCCAAAGGGGAGCGCCGACCGGG	838
Qy	753	aggactttgctgctttcgagcctggtcgctgtatggtatggtcagagcatgagctccc	812
Db	839	AGAGGCTTAGGGCCCTTCTGGAGAGGCTTGAGTTGGCAGGCTCTCTCCACGAGTTGGGCC	898
Qy	813	tgcaggaccggc	824
Db	899	TTCTGGAAGGCC	910

RESULT 8
US-08-599-491-31
; Sequence 31, Application US/08599491
; Patent No. 5846717
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; APPLICANT: GROTELUESCHEN HALL, JEFF S.
; APPLICANT: LYAMICHEV, VICTOR
; APPLICANT: OLIVE, DAVID M.
; APPLICANT: PRUDENT, JAMES R.
; TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY
; INVADER-DIRECTED CLEAVAGE

; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,491
; FILING DATE: 23-JAN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: P-40,027
; REFERENCE/DOCKET NUMBER: FORS-01802
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 948 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-599-491-31

Query Match 2.9%; Score 40.8; DB 3; Length 948;
Best Local Similarity 47.6%; Pred. No. 0.09;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy	573	ttgagaagcccgctcgtgctgagccctgagcgacacagcgccgagcccgagctga	632
Db	559	TTCTGGAGAGTGGGGAGCGCTGGGAAGCCCTCCTCAAGAACCTGGACCGGCTGAAGCCG	718
Qy	633	ccctgagccagaagataagaccagctgcagaaatccagacacctgctgagctatgtca	692
Db	719	CCATCCGGGAGAAATCTTGGCCACATGACATGACATCTCCTGGACCTCTCTGGACCTGGCC	778
Qy	693	cagtgcccaaggaagtgtccagttgggggagaggtacggtcagagagcgggagg	752
Db	779	AGGTGCGCACCGACCTGCCCTGGAGTGGACTTGGCCAAAGGGGAGCGCCGACCGGG	838
Qy	753	aggactttgctgctttcgagcctggtcgctgtatggtatggtcagagcatgagctccc	812
Db	839	AGAGGCTTAGGGCCCTTCTGGAGAGGCTTGAGTTGGCAGGCTCTCTCCACGAGTTGGGCC	898
Qy	813	tgcaggaccggc	824
Db	899	TTCTGGAAGGCC	910

RESULT 9
US-08-756-386-31
; Sequence 31, Application US/08756386
; Patent No. 5985557
; GENERAL INFORMATION:
; APPLICANT: Prudent, James R.
; APPLICANT: Hall, Jeff G.
; APPLICANT: LYAMICHEV, VICTOR I.
; TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco

ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,516
FILING DATE: 24-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01072
FILING DATE: 21-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/759,038
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,314
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/756,386
FILING DATE: 29-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 948 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
OS-08-823-516-26

Query Match	2.9%	Score 40.8;	DB 4;	Length 948;	0;
Best Local Similarity	47.6%;	pred. No. 0.09;			
Matches 120;	Conservative	0;	Mismatches 132;	Indels	Gaps
Qy 573	ttgaaagggccgcgtcctgtgagggccctgcagcagcacagcagccgagcccgagctga	632			
Db 659	TTCTGGAGGAGTGGGGAGCCCTGGGAAGCCCTCCTCAAGAACCTGGACCGGTGAAGCCCG	718			
Qy 633	ccctgagccagaagaataaggaccaagctgcagaatccagacgtctgtgagctatgtcact	692			
Db 719	CCATCCGGGAGGAAGATCTGCGCCACATGGACGATCTCAAGCTCTCTTGGGACCTGGCCA	778			
Qy 693	cagtgcccaaggaagtggctccagttggtgggggagagaggtacgggtcagagagcggggagg	752			
Db 779	AGGTGCGACCGACCTGTCCTTGAGGTGGACTTCGCCAAAAGGGGGAGCCCGACCGGG	838			
Qy 753	aggactttgtcgcctttcagccttgctgcgtgcgtatggcatgccagggcatgactccc	812			
Db 839	AGAGGCTTAGGSCCTTTCTGTGAGAGGCTTAGTTTGGCAGGCTCTCTCCACGAGTTTCGGCC	898			
Qy 813	tgcaggaccggc	824			
Db 899	TTCTGGAAGACC	910			

STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/756,386
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02564
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 948 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
OS-08-756-386-31

```

Query Match          2.9%; Score 40.8; DB 4; Length 948;
Best Local Similarity 47.6%; Pred. No. 0.09;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 573 ttgagaagcccgctgctgctctggagccctgcagcagcacagcgagcccgagactga 632
Db 659 TTCTGGAGNGTGGGGAGCCTTGGAGCCCTCCTCAGAAGACCTGGACCGGCTGAAGCCCG 718

QY 633 cctctgagccagaagataaagaccagctgcagaatccagacctgctgtagctatgctact 692
Db 719 CCATCCGGGAGAAAGATCTCTGGCCACATGGACGATCTGAAGCTCTCTCTGGGACCTGGCCA 778

QY 693 cagtgcccaagaagtggtccagttggggcgagagggctacgggtcagagagcgggaggg 752
Db 779 AGGTGCGCACCGACCTTGCCTCCCTCTGGAGGTGGACTTCGCCAAAGCGGGACCCGACCGG 838

QY 753 aggactttgctgcttctcagactggctgcgctgctatggtgatgccagggaatagctccc 812
Db 839 AGAGCCTTAGGGCCCTTCTGGAGAGGCTTGAGTTGGCAGGCTCTCTCCAGGAGTTGCGCC 898

QY 813 tgcaggaccggcc 824
Db 899 TTCTGGAAAGCC 910

RESULT 10
US-08-823-516-26
; Sequence 26, Application US/08823516
; Patent No. 5994069
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Mast, Andrea L.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: Detection of Nucleic Acids By Multiple
; TITLE OF INVENTION: Sequential Invasive Cleavages
; NUMBER OF SEQUENCES: 163
; CORRESPONDENCE ADDRESS:
;

```

[illegible]

Query Match	2.9%	Score 40.8;	DB 5;	Length 948;
Best Local Similarity	47.6%;	Pred. No. 0.09;		
Matches 120;	Conservative 0;	Mismatches 132;	Indels 0;	Gaps 0;
Qy	573	ttgagaagggccctcggtcctggagggccctgcagcagcacagccgagccggagctga	632	
Db	659	TTCTGGAGGAGTGGGGAGGCTTGAAGCCCTCTCAAGAACTGTGACCGGCTGAAGCCGG	718	
Qy	633	cccttgagccagaagaatgaagaccaggtctgagaattccagacctgctggagctatgtcact	692	
Db	719	CCATCCGGGAGAGATCCTGGCCCCACATGGACGATCTGAAGCTCTCTGGGACCTGGCCA	778	
Qy	693	cagtcgccaaggaagtgggtccagttggggggcagagactacgggtcagagagcggggaggg	752	
Db	779	AGGTGCACACGACCTGCCCTTGAGGTGGACTTCGCCAAAGCGGGGACCCGACCGGG	838	
Qy	753	aggacttgccttccttcagacctggctgcgctgctatggcatgccagggcatgactccc	812	
Db	839	AGAGGCTTTAGGGCTTTCTTGAGAGAGGCTTGAGTTTGGCAGGCTCCTCCACAGAGTTCGGCC	898	
Qy	813	tgcaggaccggcg	824	

us-09-215-435-71.rni

wed May 24 09:19:54 2000

Db 863 TTCTGGAAGCC 874

RESULT 14

US-08-483-043-11

; Sequence 11, Application US/08483043

; Patent No. 5691142

; GENERAL INFORMATION:

; APPLICANT: DAHLBERG, JAMES E.

; APPLICANT: LYAMICHEV, VICTOR I.

; APPLICANT: BROW, MARY ANN D.

; TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE DNA

; TITLE OF INVENTION: DNA POLYMERASE

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/483,043

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/073,384

; FILING DATE: 04-JUN-1993

; APPLICATION NUMBER: US 07/986,330

; FILING DATE: 07-DEC-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: CARROLL, PETER G.

; REGISTRATION NUMBER: 32,837

; REFERENCE/DOCKET NUMBER: FORS-00613

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 962 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-483-043-11

Query Match 2.9%; Score 40.8; DB 2; Length 962;

Best Local Similarity 47.6%; Pred. No. 0.091; Indels 0; Gaps 0;

Matches 120; Conservative

Qy 573 ttgagaagccctcgcttgagccctgcagcagcagccgagccgagccgagctga 632

Db 623 TTCTGGAAGAGTGGGGAGCCTCTCAAGAACCTTGACCGCTGAAGCCCG 682

Qy 633 ccctgagccagaagataagaccagctgcagaaatccagaccctgctgagctatgtact 692

Db 683 CCATCCGGGAGAAGATCCTGGCCACATGGACGATCTCTGGGACCTGGCCA 742

Qy 693 cagtgcccaagaagtggtccagttgggggagagagctacgggtcagagagcgaggag 752

Db 743 AGGTGCGCAGCAGCTGCCCTTGAGGTGGACTTCGCCAAAAGGCGGAGCCGACCGGG 802

Qy 753 aggaatttgccttttcagccttgctgctgctgctgctgctgctgctgctgctgctgct 812

Db 803 AGAGGCTTAGGGCCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCTCCACAGTTCGGCC 862

RESULT 13

US-08-254-359A-11

; Sequence 11, Application US/08254359A

; Patent No. 5614402

; GENERAL INFORMATION:

; APPLICANT: DAHLBERG, JAMES E.

; APPLICANT: LYAMICHEV, VICTOR I.

; APPLICANT: BROW, MARY ANN D.

; TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE

; TITLE OF INVENTION: DNA POLYMERASE

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/254,359A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/073,384

; FILING DATE: 06-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/986,330

; FILING DATE: 07-DEC-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: CARROLL, PETER G.

; REGISTRATION NUMBER: 32,837

; REFERENCE/DOCKET NUMBER: FORS-01000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 962 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-254-359A-11

Query Match 2.9%; Score 40.8; DB 1; Length 962;

Best Local Similarity 47.6%; Pred. No. 0.091; Indels 0; Gaps 0;

Matches 120; Conservative

Qy 573 ttgagaagccctcgcttgagccctgcagcagcagccgagccgagctga 632

Db 623 TTCTGGAAGAGTGGGGAGCCTCTCAAGAACCTTGACCGCTGAAGCCCG 682

Qy 633 ccctgagccagaagataagaccagctgcagaaatccagaccctgctgagctatgtact 692

Db 683 CCATCCGGGAGAAGATCCTGGCCACATGGACGATCTCTGGGACCTGGCCA 742

Qy 693 cagtgcccaagaagtggtccagttgggggagagagctacgggtcagagagcgaggag 752

Db 743 AGGTGCGCAGCAGCTGCCCTTGAGGTGGACTTCGCCAAAAGGCGGAGCCGACCGGG 802

Qy 753 aggaatttgccttttcagccttgctgctgctgctgctgctgctgctgctgctgctgct 812

Db 803 AGAGGCTTAGGGCCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCTCCACAGTTCGGCC 862

Qy 813 tgcaggaccggc 824

us-09-215-435-71.rni

Wed May 24 09:19:54 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:00:37 ; Search time 3550.87 Seconds
(without alignments)
-1641.394 Million cell updates/sec

Title: US-09-215-435-71

Perfect score: 1398

Sequence: 1 gatgctgagccgccgagc.....acatttgaaaaaaaaaaaaa 1398

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues

Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq.*

45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US095_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq.*
86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq.*
87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq.*
88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq.*
90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq.*
91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq.*
92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq.*
93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq.*
94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq.*
95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq.*
96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq.*
97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq.*
98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1398	100.0	1398	43	US-09-215-435-71 Sequence 71, Appl
2	1382.6	98.9	1423	76	US-60-096-116-60 Sequence 60, Appl
3	1135.2	81.2	1285	43	US-09-215-435-70 Sequence 70, Appl

c	4	1134.8	81.2	1309	76	US-60-096-116-59	Sequence 59, Appl
	5	956.8	68.4	1812	56	US-09-471-275-9883	Sequence 9883, Ap
	6	717.2	51.3	1174	43	US-09-215-435-69	Sequence 69, Appl
	7	717.2	51.3	1197	76	US-60-096-116-58	Sequence 58, Appl
	8	665.8	47.6	1899	92	US-60-172-377-10510	Sequence 10510, A
	9	514	36.8	1451	56	US-09-471-275-1784	Sequence 1784, Ap
c	10	484.8	34.7	1077	43	US-09-205-070-12701	Sequence 12701, A
c	11	484.8	34.7	1077	50	US-09-340-623-12701	Sequence 12701, A
	12	459.4	32.9	502	45	US-09-277-227-17958	Sequence 17958, A
	13	459.4	32.9	502	50	US-09-346-956-16218	Sequence 16218, A
	14	450.4	32.2	1231	52	US-09-371-168-6486	Sequence 6486, Ap
	15	434.2	31.1	519	96	US-60-185-215-3914	Sequence 3914, Ap
	16	431.8	30.9	495	49	US-09-332-782-2467	Sequence 2467, Ap
	17	427.8	30.6	432	56	US-09-489-036-32457	Sequence 32457, A
	18	420	30.0	458	45	US-09-277-227-17550	Sequence 17550, A
	19	420	30.0	458	50	US-09-346-956-15990	Sequence 15990, A
	20	404.4	28.9	1548	92	US-60-172-360-14738	Sequence 14738, A
	21	393	28.1	1458	51	US-09-353-690-6347	Sequence 6347, Ap
	22	366.4	26.2	384	56	US-09-489-036-17473	Sequence 17473, A
	23	366	26.2	783	51	US-09-354-899-7677	Sequence 7677, Ap
	24	365	26.1	542	51	US-09-354-899-3471	Sequence 3471, Ap
	25	364	26.0	509	47	US-09-293-972-6612	Sequence 6612, Ap
	26	354	25.3	456	49	US-09-332-782-977	Sequence 977, App
c	27	323.4	23.1	509	56	US-09-489-036-13428	Sequence 13428, A
	28	294.4	21.1	312	38	US-09-065-521-786	Sequence 786, App
	29	288.8	20.7	385	45	US-09-274-861-5090	Sequence 5090, Ap
	30	288.8	20.7	413	45	US-09-274-861-5089	Sequence 5089, Ap
	31	282	20.2	282	26	US-08-901-904-65	Sequence 65, Appl
	32	280.4	20.1	420	56	US-09-489-036-385	Sequence 385, App
	33	279.4	20.0	444	50	US-09-346-956-9764	Sequence 9764, App
	34	277	19.8	401	45	US-09-271-490-387	Sequence 387, App
	35	274	19.6	275	26	US-08-901-904-1384	Sequence 1384, Ap
	36	265	19.0	417	56	US-09-489-036-34133	Sequence 34133, A
	37	262	18.7	364	49	US-09-332-782-607	Sequence 607, App
	38	262	18.7	364	57	US-09-515-694-607	Sequence 607, App
	39	260	18.6	284	38	US-09-066-970-1732	Sequence 1732, Ap
	40	260	18.6	264	67	US-60-044-798-1732	Sequence 1732, Ap
	41	241.8	17.3	324	83	US-60-133-875-559	Sequence 559, App
	42	240.4	17.2	463	56	US-09-489-036-2007	Sequence 2007, Ap
c	43	239.4	17.1	375	56	US-09-489-036-15427	Sequence 15427, A
	44	236	16.9	407	56	US-09-489-036-27213	Sequence 27213, A
	45	235.4	16.8	396	47	US-09-293-972-11984	Sequence 11984, A

ALIGNMENTS

RESULT 1

US-09-215-435-71

Sequence 71, Application US/09215435A

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, Jean-Baptiste

APPLICANT: Duclert Americ,

APPLICANT: Bougueleret Lydie

TITLE OF INVENTION: Extended cDNAs

FILE REFERENCE: GENSET.019A

CURRENT APPLICATION NUMBER: US/09/215, 435A

CURRENT FILING DATE: 1998-12-17

EARLIER APPLICATION NUMBER: 60/069, 957

EARLIER FILING DATE: 1997-12-17

EARLIER APPLICATION NUMBER: 60/074, 121

EARLIER FILING DATE: 1998-2-9

EARLIER APPLICATION NUMBER: 60/081, 563

EARLIER FILING DATE: 1998-4-13

EARLIER APPLICATION NUMBER: 60/096, 116

EARLIER FILING DATE: 1998-8-10

NUMBER OF SEQ ID NOS: 519

SOFTWARE: Patent.pm

SEQ ID NO 71

LENGTH: 1398

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 2..1171

FEATURE:

NAME/KEY: sig_peptide

LOCATION: 2...205

OTHER INFORMATION: Von Heljne matrix

OTHER INFORMATION: seq 7.3

OTHER INFORMATION: seq LRLLSPLEFGAOP/QQ

FEATURE:

NAME/KEY: polyA_signal

LOCATION: 1368...1373

FEATURE:

NAME/KEY: polyA_site

LOCATION: 1386...1398

US-09-215-435-71

Query Match 100.0%; Score 1398; DB 43; Length 1398;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatgctgagggcccgagctgcacctggccagcagtttgaatgagccctgcagggc 60

Db 1 gatgctgagggcccgagctgcacctggccagcagtttgaatgagggcctgcagggc 60

QY 61 gctggtgttcggcggtcggtggaagtctctctgcagccgcaaccctgaggtgccctt 120

Db 61 gctggtgttcggcggtcggtggaagtctctctgcagccgcaaccctgaggtgccctt 120

QY 121 tgagagcagtgcctacgcacatctcagcttcagcccgcaaggagctgcgcctgatact 180

Db 121 tgagagcagtgcctacgcacatctcagcttcagcccgcaaggagctgcgcctgatact 180

QY 181 gacccctctgctggggcccgagcccaagagcagccctggtctctgcctcgcgtcgg 240

Db 181 gacccctctgctggggcccgagcccaagagcagccctggtctctgcctcgcgtcgg 240

QY 241 catgtccggtctcttttcagctggtgccccgcgagagagctgcccacacacctgcg 300

Db 241 catgtccggtctcttttcagctggtgccccgcgagagagctgcccacacacctgcg 300

QY 301 cttttacagggcccccgcctggccccggctgcgcctatgtttcgtggacaatccgcgggtt 360

Db 301 cttttacagggcccccgcctggccccggctgcgcctatgtttcgtggacaatccgcgggtt 360

QY 361 cggccctgagaccttgggggaaagtggcagccgcccggcggcctgtgtcttcgagga 420

Db 361 cggccctgagaccttgggggaaagtggcagccgcccggcggcctgtgtcttcgagga 420

QY 421 gtaccagcagttcagggagaatgtgtacgaaacctagcggataaaggccctttgaccggcc 480

Db 421 gtaccagcagttcagggagaatgtgtacgaaacctagcggataaaggccctttgaccggcc 480

QY 481 catctgcgagggccctcctggaccagaggtttcttcaatggcattggcaactatctgcgggc 540

Db 481 catctgcgagggccctcctggaccagaggtttcttcaatggcattggcaactatctgcgggc 540

QY 541 agagatcctgtaccgctgaagatccccccctttgagaagcccgcctcggtctctgcaggc 600

Db 541 agagatcctgtaccgctgaagatccccccctttgagaagcccgcctcggtctctgcaggc 600

QY 601 cctgcagcagcagcccgagcccgagcgtgcacctgagccctgagccagaataaggaccgaagct 660

Db 601 cctgcagcagcagcccgagcccgagcgtgcacctgagccctgagccagaataaggaccgaagct 660

QY 661 gcagaatccagacctgctggagctatgtcaactcagtgccccaaaggaagtggctcagttggg 720

Db 661 gcagaatccagacctgctggagctatgtcaactcagtgccccaaaggaagtggctcagttggg 720

QY 721 gggcagagggctacgggtcagagagcggggagggaggttgccttttcgagccctggct 780

Db 721 gggcagagggctacgggtcagagagcggggagggaggttgccttttcgagccctggct 780


```

; OTHER INFORMATION: region 12..521
; OTHER INFORMATION: id AA744383
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(888..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 14..512
; OTHER INFORMATION: id AA833865
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(892..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 12..504
; OTHER INFORMATION: id AA831660
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(909..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 1..478
; OTHER INFORMATION: id AA604501
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(968..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 1..419
; OTHER INFORMATION: id AA595778
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(966..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 1..421
; OTHER INFORMATION: id AA280803
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1099..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 2..289
; OTHER INFORMATION: id N34970
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1199..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 16..203
; OTHER INFORMATION: id AA927220
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1201..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 1..186
; OTHER INFORMATION: id W80730
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1010..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 19..399
; OTHER INFORMATION: id H78102
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1022..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 7..372
; OTHER INFORMATION: id AA724197
; FEATURE:

```

```

; NAME/KEY: est
; LOCATION: complement(983..1367)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 95
; OTHER INFORMATION: region 1..390
; OTHER INFORMATION: id H66509
; FEATURE:
; NAME/KEY: est
; LOCATION: 937..984
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 95
; OTHER INFORMATION: region 117..164
; OTHER INFORMATION: id AA370630
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(937..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 94
; OTHER INFORMATION: region 1..459
; OTHER INFORMATION: id R62756
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(951..1386)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 94
; OTHER INFORMATION: region 1..444
; OTHER INFORMATION: id R62756
; FEATURE:

Query Match      98.9%; Score 1382.6; DB 76; Length 1423;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1387; Conservative 11; Mismatches 0; Indels 1; Gaps 1;

QY 1 9atgctagagggccccgagctgacctgcccagccagttgtgaatgaggcctgcagggc 60
Db 1 GATGCTGAGGGCCCCGGAGCTGCACCTGGCCAGCCAGTTTGTGAATGAGGCTGCAGGCG 60
QY 61 gctggtgttcggcgctgctggtggaagaagtcctctctgacgcgcaacccctgaggtgccctt 120
Db 61 GCTGTGTGTTGGCGGCTGCGTGAGAGTCTCTGTACAGCCGCAACCCCTGAGGTGCCCTT 120
QY 121 tgagcagctgctaccgcatctcagcttcagccgcgagagctgcacgctgcgctgatact 180
Db 121 TGAGACAGTGTCTACCGCATCTCAGCTTCAGCCCGCGGCAAGGAGCTGCGCTGATACT 180
QY 181 gagccctctgctggggccccagcccaacagagagcactggccctgggttcctgcgttcgg 240
Db 181 GAGCCCTCTGCTGGGGCCCCAGCCCAACAGGAGGCACTGGCCCTGTTCTCCGTTCCG 240
QY 241 catgctcggctcttttcagctggtgccccgcgagagctgcacgctgcacccatgccactgcg 300
Db 241 CATGTCCGGCTCTTTTCAGCTGCTGCTGCCCCGGAGGAGCTGCCACGCCCATGCCACCTGGG 300
QY 301 cttttacagggccccgctggcccccggtctgcctctatgtttctgtagacatccgcggtt 360
Db 301 CTTTTACAGGGCCCCGCTGGCCCCCGCTGCGCCCTATGTTTCGTGGACATCCGCGGTT 360
QY 361 cggccgctgggaccttgggggaaaagtggcagccggcccgccctgtgctc-ttgcagg 419
Db 361 CGGCCGCTGGGACCTTGGGGGAAAGTGGCAGCGGCGGCCCTGTGTCTTTTCAGG 420
QY 420 agtaccagagttcaggagagaatgtgctacgaaacctagcgataaggcctttgacccgc 479
Db 421 AGTACCAGCAGTTTCAGGGAGAAATGCTACTAGGAACCTAGCGGATAGGCTTTGACCGGC 480
QY 480 ccatctgcgaggccctctctggaccagaggttcttctaattggcattgggaactactctcggg 539
Db 481 CCATCTGCGAGGCCCCCTCTGTGACACAGAGGTTCTTCAATGTCATTGGCAACTATCTCGGG 540
QY 540 cagagatcctgtaccggctgaagatcccccccttttagaagagcccgctcggtcctgaggg 599
Db 541 CAGAGATCCTGTACCGGCTGAAGATCCCCCCCCCTTTGAGAAAGGCCCTCTCGTCTCTGGAGG 600

```

```
Qy 600 ccctgcagcagcacagccgagcccgagctgacccctgagccagaataagagccaagc 659
Db 601 CCCTGcAGcAGcAcAGGcCGcAGcCGcAGcGTgACcCTGAGcCAGcAAGcTAAGcAGcCAGc 660
Qy 660 tgcagaatccagacccctgctgagctatgtcaactcagtgccaaagaaagtggtccagttgg 719
Db 661 TGCAGAAATCCAGACCTCTGcAGCTATGTCTACCTCAGTCCcCAAGAAAGTGcTCCAGTTGG 720
Qy 720 gggccagaggtcaggttcagagagcgggagagagactttgctccttcagccctggc 779
Db 721 GGGcCAGAGGcTACGGGTcAGAGAGcGGGAGGAGACTTTGcTTCGcCCTTCGAGcCCTGGc 780
Qy 780 tgcgtgctatggcatgccagcagtgagctccctgcaggaccggcagtcagtgccgtaccatct 839
Db 781 TGCcCTGTATGGcATGCCAGGCATAGcTCCcCTGCAGGACcCGcATGGcCGGTACCATCT 840
Qy 840 ggttcagggggataccttgagaccgttgaccaccacaaagggcgaactcccgcaaaaagaat 899
Db 841 GGTTCcAGGGGGATcCTGGACcGTGGcACCGTTGGcACCCMAAGGGSGcAAGTYcCGcCAAAAAGAAAT 900
Qy 900 ccaagccacacacagctagtcctagacacagagtggagagccttttcctccagcaagg 959
Db 901 CcAAGGGcACACAGcTAgTcCTCAGGACAGAGTGGAGGAGcCTTTGcCTCCGAGcCAAGG 960
Qy 960 ccccttcaggacacgaagggcaagagagaccttcctaaagagactgcacaaccagcggc 1019
Db 961 CCCcCTTCAGGACAGcAAGGGcCAARAGAGAcCTTCCTAAGAGGAcTGCACAcCCARCGGC 1020
Qy 1020 ctgaggggaccagcctccagcagaccacagaagctccacagtgcccaagaaaggagga 1079
Db 1021 CTkAGGGGACCAGcCTCCAGcAGGACCcAGAAAGcTCCcMcAGTGCcCAAAAGGGGAGGA 1080
Qy 1080 ggaaggggcacagagcctctggccactgcagaccgcagaccgggaaggtcaagctgacatcc 1139
Db 1081 GGAAGGGGcGACAGcGAGcCCTTTGGcCACtGCAGAcCCcCGAAAGGTCAAGGcTCAcATCC 1140
Qy 1140 catccttggaaaccagaggggacctcagcctcttagcagagaggtcctccttcttgacac 1199
Db 1141 CATcCTTGGRAcCAGAGGGGAcCTCAGcCCTTTAGcAGGAGGcCTCCTCTCTGTCACTC 1200
Qy 1200 acccttttattgtcttgccctgcattctgtgggtctgaaatttttggagcagggcaatat 1259
Db 1201 ACCcCTTTCTTAWTCTTCTTGGcCCTGCATcCTGGGGGTGTGAATTTTGGAGcAGGCAATAT 1260
Qy 1260 ctgaagtgcaaacagggccctacagctgttccctgcacaactcattgttttaattgta 1319
Db 1261 CTGAAGGTGCAAAcAGGcCCcTACGGcCTGTTCcCTGCACAAcTCTCATGGGTTTTAATTGTA 1320
Qy 1320 ccccatcttcacatctttaagctcatgtgaaaaatgctgcaatttttaataaaactgata 1379
Db 1321 CCCcATCTTCACATCTTTAAAGcTcATGTGAAAAATGCTGCATTTTAAATAACTGATA 1380
Qy 1380 catttgaaaaaataaaaa 1398
Db 1381 CATTTGAAAAAATAAAAAA 1399
```

```
RESULT 3
US-09-215-435-70
; Sequence 70, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
```

```
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 70
; LENGTH: 1285
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1051
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 2..205
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 7.3
; OTHER INFORMATION: seq LRLILSLPLFGAQP/QQ
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1248..1253
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1272..1285
US-09-215-435-70
```

```
Query Match      81.2%; Score 1135.2; DB 43; Length 1285;
Best Local Similarity 90.8%; Pred. No. 2.6e-259;
Matches 1270; Conservative 0; Mismatches 8; Indels 120; Gaps 1;
```

```
Qy 1 gatgcctgagggcccgagctgcacctggccagcagatttgtaatgagggcctgcagggc 60
Db 1 gatgcctgagggcccgagctgcacctggccagcagatttgtaatgagggcctgcagggc 60
Qy 61 gctggtgttcggcggtcgtgagaaagtcctctgcagccgaacccctgaggtgccctt 120
Db 61 gctggtgttcggcggtcgtgagaaagtcctctgcagccgaacccctgaggtgccctt 120
Qy 121 tgagagcagtgctaccacgatctcagcttcagccgcgcgaaggagctgcgctgatact 180
Db 121 tgagagcagtgctaccacgatctcagcttcagccgcgcgaaggagctgcgctgatact 180
Qy 181 gagccctcgtcgtgggcccagcccaacagagagcactggccttgcttcctcgtttgg 240
Db 181 gagccctcgtcgtgggcccagcccaacagagagcactggccttgcttcctcgtttgg 240
Qy 241 catgtccggctcttttcagctggtgccccgcgagagctgccacgccatgccacactgg 300
Db 241 catgtccggctcttttcagctggtgccccgcgagagctgccacgccatgccacactgg 300
Qy 301 cttttacagcggcccgctggccccggctgcgcctatgtttcgtggacatccgcggtt 360
Db 301 cttttacagcggcccgctggccccggctgcgcctatgtttcgtggacatccgcggtt 360
Qy 361 cggccgctgggaccttggggaaagtggcagccggcgccggcctgtgttttcaggga 420
Db 361 cggccgctgggaccttggggaaagtggcagccggcgccggcctgtgttttcaggga 420
Qy 421 gtaccagcagtttcaggaggagaaatgtgtcagaaacctagcggataaggcctttgacggcc 480
Db 421 gtaccagcagtttca ----- 435
Qy 481 catctgcgagggccctcctggaccagaggttcttcaatggttggaactatctcgggc 540
Db 435 ----- 435
Qy 541 agagatcctgtaccggctgaagatccccccctttgagaaggcccgctcggtcctggaggc 600
Db 435 -----ggctgaagatccccccctttgagaaggcccgctcggtcctggaggc 480
Qy 601 cctgcagcagcagggccgagcccgagctaccctgagcagagataagaccacgaact 660
Db 601 cctgcagcagcagggccgagcccgagctaccctgagcagagataagaccacgaact 660
```

```
Db 481 cctgcagcagcagcagcccgagcccgagctgacccctgagccagaagataagaccagct 540
Qy 661 qcagaatccagacctgctgagctatgtcactcagtgcccaagggaagtgtccagttggg 720
Db 541 qcagaatccagacctgctgagctatgtcactcagtgcccaagggaagtgtccagttggg 600
Qy 721 ggcagaggtcagcgtcagagagcgggagagagacttgcctcttcgagcctggct 780
Db 601 ggcaggggctacgggtcagagagcgggagagagacttgcctcttcgagcctggct 660
Qy 781 gcctgtatggtcagcagcagcagcctccctgcaggaccggcagtgccgtaccatctg 840
Db 661 gcctgtatggtcagcagcagcagcctccctgcaggaccggcagtgccgtaccatctg 720
Qy 841 gttccagggggtcctgacccgttggcaccacaaaggcgcgaagtcctccgcacaaagaatc 900
Db 721 gttccagggggtcctgacccgttggcaccacaaaggcgcgaagtcctccgcacaaagaatc 780
Qy 901 caagggccacacagctgagctcagacagagtgagagcgttttgcctccgagcaagc 960
Db 781 caagggccacacagctgagctcagacagagtgagagcgttttgcctccgagcaagc 840
Qy 961 cccctccaggacacgaaggcgaagagagaccttcttaagagagactgcacccagcgcc 1020
Db 841 cccctccaggacacgaaggcgaagagagaccttcttaagagagactgcacccagcgcc 900
Qy 1021 tgaggggaccagcctccagcagaccagagctccacagagctcccaagaaggaggag 1080
Db 901 tgaggggaccagcctccagcagaccagagctcccaagaaggaggag 960
Qy 1081 gaaggggacagcagcctctgcccactgcagaccccggaaggtcaagctgacatccc 1140
Db 961 gaaggggacagcagcctctgcccactgcagaccccggaaggtcaagctgacatccc 1020
Qy 1141 atccttgaaacagagggagcctcagcctcttagcagagagctccttgcctgcactca 1200
Db 1021 atccttgaaacagagggagcctcagcctcttagcagagagctccttgcctgcactca 1080
Qy 1201 cccctttattgtctgcccagcagctgctgagagagagctccttgcctgcactca 1260
Db 1081 cccctttattgtctgcccagcagctgctgagagagagctccttgcctgcactca 1140
Qy 1261 tgaaggtcaaacagcgcctcagcgtgtccctgcacacactctcatgttttaattgtac 1320
Db 1141 tgaaggtcaaacagcgcctcagcgtgtccctgcacacactctcatgttttaattgtac 1200
Qy 1321 cccatctccacatctttaaagctcagtgtaaaatgctgcaatttttaaaactgatac 1380
Db 1201 cccatctccacatctttaaagctcagtgtaaaatgctgcaatttttaaaactgatac 1260
Qy 1381 atttgaataaaataaaaaa 1398
Db 1261 atttgaataaaataaaaaa 1278

RESULT 4
US-60-096-116-59
; Sequence 59, Application US/60096116
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/60/096.116
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: GENSET.038PR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 1309 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Lymphocytes ganglia
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 2..205
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 7.3
OTHER INFORMATION: seq LRLILSLPLGAQP/QQ
FEATURE:
NAME/KEY: poly_a_signal
LOCATION: 1248..1253
FEATURE:
NAME/KEY: poly_a
LOCATION: 1272..1285
IDENTIFICATION METHOD: blastn2
FEATURE:
NAME/KEY: est
LOCATION: complement(1149..1267)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 10..128
OTHER INFORMATION: id AA731925
FEATURE:
NAME/KEY: est
LOCATION: complement(847..1265)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..419
OTHER INFORMATION: id AA595778
FEATURE:
NAME/KEY: est
LOCATION: complement(909..1267)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 39..396
OTHER INFORMATION: id AA824509
FEATURE:
NAME/KEY: est
LOCATION: complement(912..1267)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 33..388
OTHER INFORMATION: id AA832090
FEATURE:
NAME/KEY: est
LOCATION: complement(985..1267)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 24..306
OTHER INFORMATION: id AA688993
FEATURE:
NAME/KEY: est
```

```

1 / LOCATION: complement(756..1267)
2 IDENTIFICATION METHOD: blastn2
3 OTHER INFORMATION: identity 98
4 OTHER INFORMATION: region 10..521
5 OTHER INFORMATION: id AA744383
6 FEATURE:
7 NAME/KEY: est
8 LOCATION: complement(767..1267)
9 IDENTIFICATION METHOD: blastn2
10 OTHER INFORMATION: identity 98
11 OTHER INFORMATION: region 12..512
12 OTHER INFORMATION: id AA833865
13 FEATURE:
14 NAME/KEY: est
15 LOCATION: complement(771..1267)
16 IDENTIFICATION METHOD: blastn2
17 OTHER INFORMATION: identity 98
18 OTHER INFORMATION: region 10..504
19 OTHER INFORMATION: id AA831660
20 FEATURE:
21 NAME/KEY: est
22 LOCATION: complement(788..1265)
23 IDENTIFICATION METHOD: blastn2
24 OTHER INFORMATION: identity 98
25 OTHER INFORMATION: region 1..478
26 OTHER INFORMATION: id AA604501
27 FEATURE:
28 NAME/KEY: est
29 LOCATION: complement(845..1265)
30 IDENTIFICATION METHOD: blastn2
31 OTHER INFORMATION: identity 98
32 OTHER INFORMATION: region 1..421
33 OTHER INFORMATION: id AA280803
34 FEATURE:
35 NAME/KEY: est
36 LOCATION: complement(901..1267)
37 IDENTIFICATION METHOD: blastn2
38 OTHER INFORMATION: identity 98
39 OTHER INFORMATION: region 5..372
40 OTHER INFORMATION: id AA724197
41 FEATURE:
42 NAME/KEY: est
43 LOCATION: complement(978..1266)
44 IDENTIFICATION METHOD: blastn2
45 OTHER INFORMATION: identity 98
46 OTHER INFORMATION: region 1..289
47 OTHER INFORMATION: id N34970
48 FEATURE:
49 NAME/KEY: est
50 LOCATION: complement(985..1271)
51 IDENTIFICATION METHOD: blastn2
52 OTHER INFORMATION: identity 98
53 OTHER INFORMATION: region 18..304
54 OTHER INFORMATION: id AA883584
55 FEATURE:
56 NAME/KEY: est
57 LOCATION: complement(1078..1267)
58 IDENTIFICATION METHOD: blastn2
59 OTHER INFORMATION: identity 98
60 OTHER INFORMATION: region 14..203
61 OTHER INFORMATION: id AA927220
62 FEATURE:
63 NAME/KEY: est
64 LOCATION: complement(1080..1265)
65 IDENTIFICATION METHOD: blastn2
66 OTHER INFORMATION: identity 98
67 OTHER INFORMATION: region 1..186
68 OTHER INFORMATION: id W80730
69 FEATURE:
70 NAME/KEY: est
71 LOCATION: complement(889..1271)
72 IDENTIFICATION METHOD: blastn2
73 OTHER INFORMATION: identity 97

```

```

; OTHER INFORMATION: region 13..399
; OTHER INFORMATION: id H78102
; FEATURE:
; NAME/KEY: est
; LOCATION: 599..712
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 30..143
; OTHER INFORMATION: id AA359490
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(816..1265)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 1..459
; OTHER INFORMATION: id R62756
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(830..1265)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 1..444
; OTHER INFORMATION: id R62756
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(862..1246)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 1..390
; OTHER INFORMATION: id H66509
; FEATURE:
; NAME/KEY: est
; LOCATION: 784..1102
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 1..327
; OTHER INFORMATION: id R62755
; FEATURE:

```

Query Match 81.2%; Score 1134.8; DB 76; Length 1309;

Query match	61.2%;	Score	1134.6;	DB
Best Local Similarity	90.88;	Pred. No.	3.3e-259;	

Best local similarity	Freq. NO.	Indels	Gaps
Matches 1269; Conservative	1; Mismatches 8;	120;	1;
Matches 1269; Indels	1; Mismatches 8;	120;	1;
Matches 1269; Gaps	1; Mismatches 8;	120;	1;

[illegible]

Db 421 GTACCAGCAGTTCA----- 435
Qy 481 catctcgaggccctccttgaccagaggtttcttaattgcatggcaactatctcgggc 540
Db 435 ----- 435
Qy 541 agagatcctgtaccgggtgaagatccccccctttgagaagggcccgctcggtcgagagc 600
Db 435 -----GCTGAAGTCCCCCTTTGAGAAGGCCCTCGGTCTGGAGC 480
Qy 601 cctgcagcagcagccgagcccgagctgacctgacctgagccagagataagaccgaact 660
Db 481 CCTGCAGCAGCAGCCGAGCCGAGCGAGCTGACCTCTGAGCCAGAGATAAGACCAAGCT 540
Qy 661 gcagaatccagacctgtgagctatgtcaactcagtgcccaaggaagtgtccagattgg 720
Db 541 GCAGAATCCAGACCTCTGGAGCTATGTCACTCAGTGCACCAAGGAAGTGAGCAGTTGGG 600
Qy 721 gggcagaggtacgggtcagagagcggggagggaggttctgaccttttcgacctggct 780
Db 601 GGCAGGGGCTACGGGTCAGAGAGCGGGAGGAGGACTTTGCTGCCTTTTCGAGCCTGGCT 660
Qy 781 ggcgtgtatggcatgccagggatgagctccctgcagggaccggcagtcgacctaccatctg 840
Db 661 GCCTGTGATGGCATGCCAGGCATGAGCTCCCTGCAAGACCCGGCATGGCCGTACCATCTG 720
Qy 841 gttccaggggatacctggaccgttgccaccacaaaggcgcaagtcccgcaaaagaaatc 900
Db 721 GTTCCAGGGGGATCTCTGGACCTTTGGACCCCAAGGGCGCAAGTCCCGCAAAAGAAATC 780
Qy 901 caagggcacacagctgagtcctgagacagagtgaggagcgttttccctccgagcaagc 960
Db 781 CAAGGCCACACAGCTGAGTCTTGAGGACAGAGTGGAGGAGCTTTGCCCTCCAAGCAAGC 840
Qy 961 ccttcacagacacagggcaagagagaccttcttaagaggagctgcaaccagcggcc 1020
Db 841 CCCTTCAAGACACAGAGGGCAAGAGAGAGACCTTCTTAAGAGGACTGCAACCCARCGGC 900
Qy 1021 tgaagggacagcctccagcagcagcagagctccacagtgcccaagaggaggag 1080
Db 901 TGAGGGACACAGCTCCAGCAGCAGCAGAGCTCCACAGTCCCAAGAGGGAGGAG 960
Qy 1081 gaaggggacagcagcctctggccactgagccactgagaccccggaaggtcaaggtgacatcc 1140
Db 961 GAAGGGCGCAGCAGCAGCTCTGCCACTGCAGACCCCGGAAGGTCAAGGCTGACATCCC 1020
Qy 1141 atccttgaaccagaggggacctcagcctcttagcagagagctcctctgttgcactca 1200
Db 1021 ATCCTTGAACCAAGAGGGAGCCTCAGCCTCTTAGCAGGAGGCTCTCCTTGTTCGACTCA 1080
Qy 1201 cctttcttattgtctgacctgcatctgggggtctgaaatttttgggagcaggaatc 1260
Db 1081 CCCTTTCTTATTGCTTGGCTGCTGCTGCTGGGCTCTGAATTTTGGGAGCAGCAATATC 1140
Qy 1261 tgaaggtgcaaacagggccctacgggtgttccctgcacaaactctcatggttttaattgtac 1320
Db 1141 TGAAGGTGCAAAACAGGCCCTACGGCTTCTCCCTGCACAACTCTCATGGTTTAAATGTAC 1200
Qy 1321 cccatcttccacatctttaagctcatgtgaaataatgtcatttttaataaactatc 1380
Db 1201 CCCATCTTCCACATCTTTAAAGCTCATGTGAAATAATGCTGCATTTTTTAATAAATGATAC 1260
Qy 1381 atttgaataaaaaaaa 1398
Db 1261 ATTTGAATCCAAAAA 1278

RESULT 5

US-09-471-275-9883
; Sequence 9883, Application US/09471275
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Contigs Obtained

; TITLE OF INVENTION: From Various Libraries
; FILE REFERENCE: 782
; CURRENT APPLICATION NUMBER: US/09/471,275
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/235,076
; EARLIER FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: US 09/234,611
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: US 09/240,371
; EARLIER FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: US 09/277,227
; EARLIER FILING DATE: 1999-03-25
; EARLIER APPLICATION NUMBER: US 09/271,490
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: US 09/293,972
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: US 09/274,861
; EARLIER FILING DATE: 1999-03-23
; EARLIER APPLICATION NUMBER: US 60/125,453
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: US 60/126,605
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: US 09/306,350
; EARLIER FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 09/399,720
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER: US 09/404,284
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER:
; EARLIER FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 10451
; SOFTWARE: pt_ct_genes Version 1.0
; SEQ ID NO 9883
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1721)..(0)
; OTHER INFORMATION: similar to gi214042 in the genepept database release 114,
; OTHER INFORMATION: Run with FASTX 3.3t00, default parameters
US-09-471-275-9883

Query Match 68.4%; Score 956.8; DB 56; Length 1812;
Best Local Similarity 99.8%; Pred. No. 6.5e-217;
Matches 958; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gatgcctgagggcccgagctgacctggccagctgttgaaatgagggcctgcagggc 60
Db 853 gatgcctgagggcccgagctgacctggccagctgttgaaatgagggcctgcagggc 912
Qy 61 gctggtgttcgagcgtcgctggagagtcctctgtagccgcaaacctgaggtgcccctt 120
Db 913 gctggtgttcgagcgtcgctggagagtcctctgtagccgcaaacctgaggtgcccctt 972
Qy 121 tgagagcagtgctaccgcatctcagcttcagcccgggcaggaagctgcgctgatact 180
Db 973 tgagagcagtgctaccgcatctcagcttcagcccgggcaggaagctgcgctgatact 1032
Qy 181 gagccctctgctggggccagcccaacagagccactggccctggtcttccgcttcgg 240
Db 1033 gagccctctgctggggccagcccaacagagccactggccctggtcttccgcttcgg 1092
Qy 241 catgtccgctcttttcagctggtgccccgcgagggagctgcccacgcatgccacactcg 300
Db 1093 catgtccgctcttttcagctggtgccccgcgagggagctgcccacgcatgccacactcg 1152
Qy 301 cttttacagggccccgcctggccccgcctgccccctatgttttctgtagacatccgcgggt 360
Db 1153 cttttacagggccccgcctggccccgcctgccccctatgttttctgtagacatccgcgggt 1212
Qy 361 cggccgctgggaccttgggggaaagtggcagccggcccgccctgtgtcttgacagga 420

|||||
Db 1213 cggcgctgggaccttgggaaagtggcagccggcgcgccgtgtcttgacgga 1272
Qy 421 gtaccagcagttcaggagaatgtgtctacgaaacctagcgataagccctttgaccggcc 480
Db 1273 gtaccagcagttcaggagaatgtgtctacgaaacctagcgataagccctttgaccggcc 1332
Qy 481 catctgagggccctctctgaccagaggttcttcaatggcattggcaactatctgcgggc 540
Db 1333 catctgagggccctctctgaccagaggttcttcaatggcattggcaactatctgcgggc 1392
Qy 541 agagatcctgtaccggctgaagatccccccctttgagaaggcccgctcggtccttggaggc 600
Db 1393 agagatcctgtaccggctgaagatccccccctttgagaaggcccgctcggtccttggaggc 1452
Qy 601 cctgcagcagcacaggccgagccggagctgacccctgagccagagataaggaaccaagct 660
Db 1453 cctgcagcagcacaggccgagccggagctgacccctgagccagagataaggaaccaagct 1512
Qy 661 gcagaatccagacctgtgagctatgtcactcagtgcccaaggaaagtgtccagttggg 720
Db 1513 gcagaatccagacctgtgagctatgtcactcagtgcccaaggaaagtgtccagttggg 1572
Qy 721 gggcagaggctacgggtcagagagcggggaggaggaactttgtcctcttcgagcctggct 780
Db 1573 gggcagaggctacgggtcagagagcggggaggaggaactttgtcctcttcgagcctggct 1632
Qy 781 gcgtgtatggaatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
Db 1633 gcgtgtatggaatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1692
Qy 841 gttccaggggagctcggagcgttggcaccagagggcgcaagtcgccgcaaaagaaatc 900
Db 1693 gttccaggggagctcggagcgttggcaccagagggcgcaagtcgccgcaaaagaaatc 1752
Qy 901 caaggccacacagctgagctcgtgagcagagtgaggagcagcttgcctcgcagcaaggc 960
Db 1753 caaggccacacagctgagctcgtgagcagagtgaggagcagcttgcctcgcagcaaggc 1812

RESULT 6

US-09-215-435-69
; Sequence 69, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 69
; LENGTH: 1174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..757
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 2..205
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 7.3

; OTHER INFORMATION: seq LRLILSLPLGAQP/QQ
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1160..1174
US-09-215-435-69

Query Match 51.3%; Score 717.2; DB 43; Length 1174;
Best Local Similarity 99.6%; Pred. No. 4.2e-160;
Matches 719; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gatccctgagggccccagctgcacctggccagccagtttgtgaatgaggcctgcagggc 60
Db 1 gatccctgagggccccagctgcacctggccagccagtttgtgaatgaggcctgcagggc 60
Qy 61 gctggtgttcggcggtcgctggagaaagtcctctgtacgcccgaacccctgagggtccctt 120
Db 61 gctggtgttcggcggtcgctggagaaagtcctctgtacgcccgaacccctgagggtccctt 120
Qy 121 tgagagcagtgccctaccgcacatctcagcttcagcccgccgagcagctgcgcctgatact 180
Db 121 tgagagcagtgccctaccgcacatctcagcttcagcccgccgagcagctgcgcctgatact 180
Qy 181 gagccctctgctggggccccagcccaacaggagccactggccctgttcttcgcgttcgg 240
Db 181 gagccctctgctggggccccagcccaacaggagccactggccctgttcttcgcgttcgg 240
Qy 241 catctcgggctctttcagctggtgccccgcgagagctgcacgcccattgcccactgcg 300
Db 241 catctcgggctctttcagctggtgccccgcgagagctgcacgcccattgcccactgcg 300
Qy 301 cttttacagcccccgcctgccccggctcgccctagtccctatgtttctgtggacatcccggtt 360
Db 301 cttttacagcccccgcctgccccggctcgccctagtccctatgtttctgtggacatcccggtt 360
Qy 361 cggcgctgggaccttggggaaagtggcagccggcgccggccctgtgtcttgacgga 420
Db 361 cggcgctgggaccttggggaaagtggcagccggcgccggccctgtgtcttgacgga 420
Qy 421 gtaccagcagttcaggagaatgtgtctacgaaacctagcgataagcccttgaccggcc 480
Db 421 gtaccagcagttcaggagaatgtgtctacgaaacctagcgataagcccttgaccggcc 480
Qy 481 catctgcagccctctctgaccagaggttcttcaatggcattggcaactatctgcgggc 540
Db 481 catctgcagccctctctgaccagaggttcttcaatggcattggcaactatctgcgggc 540
Qy 541 agagatcctgtaccggctgaagatccccccctttgagaaggcccgctcggtccttggaggc 600
Db 541 agagatcctgtaccggctgaagatccccccctttgagaaggcccgctcggtccttggaggc 600
Qy 601 cctgcagcagcacaggccgagccggagctgacccctgagccagcagataaggaaccaagct 660
Db 601 cctgcagcagcacaggccgagccggagctgacccctgagccagcagataaggaaccaagct 660
Qy 661 gcagaatccagacctgtgagctatgtcactcagtgcccaaggaaagtgtccagttggg 720
Db 661 gcagaatccagacctgtgagctatgtcactcagtgcccaaggaaagtgtccagttggg 720
Qy 721 gg 722
Db 721 tg 722

RESULT 7

US-60-096-116-58
; Sequence 58, Application US/60096116
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 501 West Broadway
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92101-3505
 COMPUTER READABLE FORM:
 MEDIUM TYPE: IBM PC compatible
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Win95
 SOFTWARE: Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/60/096,116
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Israel, Ned A.
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: GENSET.038PR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 58:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1197 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: DOUBLE
 TOPOLOGY: LINEAR
 MOLECULE TYPE: CDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo Sapiens
 TISSUE TYPE: Lymphocytes ganglia
 FEATURE:
 NAME/KEY: sig_peptide
 LOCATION: 2..205
 IDENTIFICATION METHOD: Von Heijne matrix
 OTHER INFORMATION: score 7.3
 OTHER INFORMATION: seq LRLLSLPGAQP/QQ
 FEATURE:
 NAME/KEY: poly_a
 LOCATION: 1160..1173
 IDENTIFICATION METHOD: blastn2
 US-60-096-116-58

Query Match 51.3%; Score 717.2; DB 76; Length 1197;
 Best Local Similarity 99.6%; Pred. No. 4.2e-160;
 Matches 719; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gatgctagggccgagctgacacgagccagctgtgtgaatgagggcctgcagggc 60
 DB 1 GATGCTGAGGGCCCGGAGCTGACCTGCGCAGCCAGCTTGTGATGAGGCGCTGCAGGGC 60
 QY 61 gctggtgttcggcggtgctggtgagaaagctctctgtcagccgcaacccctgaggtgcctt 120
 DB 61 GCTGgtgttcggcggtgctggtgagaaagctctctgtcagccgcaacccctgaggtgcctt 120
 QY 121 tsagagcagtgctaccgcatctcagcttcagccgagcagagagctgcgctgatact 180
 DB 121 TGAGAGCAGTGCTACCGCATCTCAGCTTCAGCCCGCGGAGAGAGCTGCGCTGTACT 180
 QY 181 gagccctctgctggggccgagcccaacagagagcactggccctgtgtcttcocgtcgg 240
 DB 181 GAGCCCTCTGCTGGGGCCCGAGCCCTCAACAGAGGACCACTGGCCCTGTGTCTTCGGCTTCGG 240
 QY 241 catgtccggctcttttcagctggtgccccgcgagagagctgcaacgcatgccacctgcg 300
 DB 241 CATGTCCGGCTCTTTTTCAGCTGTGTGTCGCGGAGAGTGGCCACGCCATGCCACTCGC 300
 QY 301 cttttacagggcccgccctggccccggtcgcctctatgtttctgtggacatccgcggtt 360
 DB 301 CTTTtACAGGGCCCGCCCTGGCCCCGGCTCGCCCTATGTTCGTGGACATCCGCGCGGT 360

QY 361 cgccgctgggaccttggggaaagtgcagccggggccggggccctgtgtgtcttcagga 420
 DB 361 CGCCGCTGGGACCTTGGGGAAAGTGCAGCGGGGCCGGCCCTGTCTTTCAGGA 420
 QY 421 gtaccagagttcaggagagatgtctacaaaacctagcggataaaggcctttaccggcc 480
 DB 421 GTACCAGAGTTCAGGGAGAGATGTCTACGAAACCTAGCGGATAAGGCCCTTTCACCGGCC 480
 QY 481 catctggaggccctctgaccagaggttctcaatggcattgtgcaactatcttcggggc 540
 DB 481 CATCTGCGAGGCCCTCTCTGGACAGAGGTTCTCAATGGCATTTGGCAACTATCTGCGGGC 540
 QY 541 agatcctgtaccgggtgaagatccccccctttgagaagggcccgctcgtcttcggggc 600
 DB 541 AGAGATCTGTACCGGCTGAAGATCCCCCCTTTGAGAAGGCCGCTCGGTCTTGGAGGC 600
 QY 601 cctgcagcagcacaggccgagcccgagctgacctgagccagagataagagaccaagct 660
 DB 601 CCTGCAGCAGCACAGGCCGAGCCCGGAGCTGACCTTGAGCCAGAGATAGGACCAAGCT 660
 QY 661 gcagaatccagacctgctggagctatgtcactcagtgcccccaaggaagtgtccagttggg 720
 DB 661 GCAGAAATTCAGACCTGCTGGAGCTATGTCACTAGTGCCCAAGGAAGTGTCCAGTTGGG 720
 QY 721 gg 722
 DB 721 TG 722
 RESULT 8
 US-60-172-377-10510
 ; Sequence 10510, Application US/60172377
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, MacDonald
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Dlep, Dinh
 ; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
 ; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym
 ; FILE REFERENCE: GX-0008 P
 ; CURRENT APPLICATION NUMBER: US/60/172,377
 ; CURRENT FILING DATE: 1999-12-16
 ; NUMBER OF SEQ ID NOS: 13,749
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 10510
 ; LENGTH: 899
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: Incyte ID No: 903209.3
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: 4, 10, 794, 856, 863, 868
 ; OTHER INFORMATION: a, t, c, g, or other
 ; US-60-172-377-10510
 Query Match 47.6%; Score 665.8; DB 92; Length 899;
 Best Local Similarity 99.0%; Pred. No. 6.1e-148;
 Matches 678; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
 QY 431 ttacaggagagatgtctacgaaacctagcggataaaggcctttgaccggcccatctgcgag 490
 DB 216 tcacaggagagatgtctacgaaacctagcggataaaggcctttgaccggcccatctgcgag 275
 QY 491 gccctcctgaccagaggttctcaatggcattgtgcaactatctgcggcgagagatccctg 550
 DB 276 gccctcctgaccagaggttctcaatggcattgtgcaactatctgcggcgagagatccctg 335
 QY 551 taccgggtgaagatccccccctttgagaagggcccgctcgtcgtcctgcagccctgcagcag 610
 DB 336 taccgggtgaagatccccccctttgagaagggcccgctcgtcgtcctgcagccctgcagcag 395


```
Qy 611 cacaggccgagccgagctgacctgagccagagataagaccagctcgaatacca 670
Dy 396 cacaggccgagccgagctgacctgagccagagataagaccagctcgaatacca 455
Qy 671 gacctgctggagctatgtcaactcagtgcccaagaagtgtccagttggggcagagcc 730
Dy 456 gacctgctggagctatgtcaactcagtgcccaagaagtgtccagttggggcagagcc 515
Qy 731 tacgggtcagagagcgggagagactttgctcccttcgagctgagctgcgtgctat 790
Dy 516 tacgggtcagagagcgggagagactttgctcccttcgagctgagctgcgtgctat 575
Qy 791 ggcagtcgagcagctgagctccctgcagaccgagctgagccgctaccatctggtccaggg 850
Dy 576 ggcagtcgagcagctgagctccctgcagaccgagctgagccgctaccatctggtccaggg 635
Qy 851 gacctgagccgttggcaccacaaagggcgcaagtcccgcaaaagaaatccaaagccaca 910
Dy 636 gacctgagccgttggcaccacaaagggcgcaagtcccgcaaaagaaatccaaagccaca 695
Qy 911 cagctgagctcagagcagagtagagacgctttgctcccgagcaagcccttcagg 970
Dy 696 cagctgagctcagagcagagtagagacgctttgctcccgagcaagcccttcagg 755
Qy 971 acacgaaggcaagagagacttctctaaagagctgcaaccagcgcctgaggggacc 1030
Dy 756 acacgaaggcaagagagacttctctaaagagctgcaaccagcgcctgaggggacc 815
Qy 1031 agctccagcagggaccacgaagctccacagtgcccaagaaggaggaagggcgga 1090
Dy 816 agctccagcagggaccacgaagctccacagtgcccaagaaggaggaagggcgga 874
Qy 1091 caggcagcctctggccactcgagac 1115
Dy 875 caggcagcctctggccactcgagac 899
```

RESULT 9

```
US-09-471-275-1784/c
; Sequence 1784, Application US/09471275
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Contigs Obtained
; FILE REFERENCE: 782
; CURRENT APPLICATION NUMBER: US/09/471,275
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/235,076
; EARLIER FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: US 09/234,611
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: US 09/240,371
; EARLIER FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: US 09/277,227
; EARLIER FILING DATE: 1999-03-25
; EARLIER APPLICATION NUMBER: US 09/271,490
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: US 09/293,972
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: US 09/274,861
; EARLIER FILING DATE: 1999-03-23
; EARLIER APPLICATION NUMBER: US 60/125,453
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: US 60/126,605
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: US 09/306,350
; EARLIER FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 09/399,720
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER: US 09/404,284
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER:
; EARLIER FILING DATE: 1999-12-16
```

```
; NUMBER OF SEQ ID NOS: 10451
; SOFTWARE: PT_CT_genes Version 1.0
; SEQ ID NO 1784
; LENGTH: 1451
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (535)...(116)
; OTHER INFORMATION: similar to gi552075 in the genepept database release 114,
; OTHER INFORMATION: Run with FASTXY 3.3t00, default parameters
US-09-471-275-1784
```

```
Query Match 36.8%; Score 514; DB 56; Length 1451;
Best Local Similarity 98.1%; Pred. No. 7.7e-112;
Matches 520; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Qy 869 cccaaaggcgcaagtcgccgcaaaagaaatccaaaggccacacagctgagctcctgagac 928
Dy 535 CCTGCAGGGCGCAAGTCCCGCAAAAGAAATCCAAAGGCCACACAGCTGAGTCTCTGAGGAC 476
Qy 929 agagtggagagcgtcttccctccgagcaagggccctccagggacacgaagggcaagaga 988
Dy 475 AGAGTGGAGGAGCGCTTTCCTCCAAAGCAAGGCCCTTCAGGACACGAAGGCAAGAGAGA 416
Qy 989 gaccttcttaagagactgcaaccagcggcctgaggggaccagctccagcaggaccaca 1048
Dy 415 GACCTTCTTAAGAGAGACTGCAACCCAGCGGCTGAGGGGACCAGCTCCAGCAGGACCCA 356
Qy 1049 gaagctccacagtcgcccaagaaggagggaggaagggcgacagggcgagcctctggccac 1108
Dy 355 GAAGCTCCACACAGTCCCAAGNAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 296
Qy 1109 tgcagaccgggagaggtcaaggtgacatcccatccttggaaaccagaggggacctcagcc 1168
Dy 295 TGCAGACCCGGGAGGTCAAGGCTGACATCCCATCCTTTGGAAACCAGAGGGGACCTCAGCC 236
Qy 1169 tcttagcaggaggtctctctgtcttgctgcaactcaaccttctattgttctgacctcatet 1228
Dy 235 TCTTAGCAGGAGGCTCTCCTTGTGTCACCTCACCTTTCTTATTGCTGTGCTGCTCATCT 176
Qy 1229 ggggtctgaatttttggagcaggaataatctgaagtgcaaacagggccctcacggctgt 1288
Dy 175 GGGGTCTGAATTTTGGGAGCAGGCAATATCTGAAGGTGCAACAGGCCCTTACGGCTGT 116
Qy 1289 tccctgcacactctcatgtgttttaattgtaccatcttccacatctttaaagctcatg 1348
Dy 115 TCCCTGCACAACTCTCATGGTTTAAATTGATACCCCATCTTCCACATCTTTAAAGCTCATG 56
Qy 1349 tgaataatgctgcatcttttaataaactgatacatattgaaaaaataaaaaa 1398
Dy 55 TGAAAAATGCTGCATTTTAAATAATTTGATACATTTTGAACCTTCAAAAAA 6
```

RESULT 10

```
US-09-205-070-12701/c
; Sequence 12701, Application US/09205070
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-748
; CURRENT APPLICATION NUMBER: US/09/205,070
; CURRENT FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12701
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-205-070-12701
```


Wed May 24 09:19:54 2000

us-09-215-435-71.rnp

QY 421 gtaccagcagttcaggagga 440
| | | | | | | | | | | | | | | | | |
Db 487 gtaccagcagttcagaaaca 506

Search completed: May 24, 2000, 01:01:00
Job time: 14232 sec

QY 1356 tgctgaatttttaataaactgatacttgaa 1387
| | | | | | | | | | | | | | | | | |
Db 1186 tgctgaatttttaataaactgatacttgaa 1217

RESULT 15
US-60-185-215-3914
; Sequence 3914, Application US/60185215
; GENERAL INFORMATION:
; APPLICANT: Hodgson, David M. Jones, Anissa L.
; APPLICANT: Lincoln, Stephen E. Yu, Jimmy Y.
; APPLICANT: Russo, Frank D. Greenawalt, Lilla B.
; APPLICANT: Spiro, Peter A. Panzer, Scott R.
; APPLICANT: Banville, Steve C. Roseberry, Ann M.
; APPLICANT: Bratcher, Shawn R. Wright, Rachel J.
; APPLICANT: Dufour, Gerard E. Chen, Wensheng
; APPLICANT: Cohen, Howard J. Liu, Tommy
; APPLICANT: Rosen, Bruce Yap, Pierre E.
; APPLICANT: Shah, Purvi Stockdreher, Theresa K.
; APPLICANT: Chalup, Michael S. Amshey, Stefan
; APPLICANT: Hillman, Jennifer L. Fong, Willy Tuen
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-0134 P
; CURRENT APPLICATION NUMBER: US/60/185,215
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 4900
; SOFTWARE: PERL Program
; SEQ ID NO 3914
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 332652.1.1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 392
; OTHER INFORMATION: a, t, c, g, or other
US-60-185-215-3914

Query Match 31.1%; Score 434.2; DB 96; Length 519;
Best Local Similarity 99.1%; Pred. No. 4.7e-93;
Matches 436; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 gatgctgagggcccgagctgcacgtgcccagcagtttgtaatgagggcctgcagggc 60
| | | | | | | | | | | | | | | | | |
Db 67 gatgctgagggcccgagctgcacgtgcccagcagtttgtaatgagggcctgcagggc 126
QY 61 gctgggttcggcggtcgctgaggaagtcctctgtcagccgcaacctgaggtgacctt 120
| | | | | | | | | | | | | | | | | |
Db 127 gctgggttcggcggtcgctgaggaagtcctctgtcagccgcaacctgaggtgacctt 186
QY 121 tgagagcagtgctctacgcacgtcagcttcagcccgccgaggaaggctgcgctgatact 180
| | | | | | | | | | | | | | | | | |
Db 187 tgagagcagtgctctacgcacgtcagcttcagcccgccgaggaaggctgcgctgatact 246
QY 181 gagccctctgcctggggccagcccaacagagccactggccctggcttcctcgttcgg 240
| | | | | | | | | | | | | | | | | |
Db 247 gagccctctgcctggggccagcccaacagagccactggccctggcttcctcgttcgg 306
QY 241 catgtccggtctcttcagctggtgcccgcgagggagctgccacgcatgccacactgcg 300
| | | | | | | | | | | | | | | | | |
Db 307 catgtccggtctcttcagctggtgcccgcgagggagctgccacgcatgccacactgcg 366
QY 301 cttttacacggcccgccctggcccccgctgcacctatgtttcgttgacatccgcggtt 360
| | | | | | | | | | | | | | | | | |
Db 367 cttttacacggcccgccctggcccccgctgcacctatgtttcgttgacatccgcggtt 426
QY 361 cgcccgctgggacattgggaaagtggcagccgcccggggccctgtgtcttcagga 420
| | | | | | | | | | | | | | | | | |
Db 427 cggccgctgggacattgggaaagtggcagccgcccggggccctgtgtcttcagga 486

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:03:02 ; Search time 3267.04 Seconds
(without alignments)
1734.416 Million cell updates/sec

Title: US-09-215-435-71
Perfect score: 1398
Sequence: 1 gatgcctgagggcccgagc.....acattgaaaaaaaaa 1398

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query
--------	---	-------

Qy 1356 tgcgcattttaaaactgatacatttgaaaaa 1396
 Db 41 TGCTGCATTTTAAATAACTGATACATTGAAAAA 1

RESULT 2
 AA833865/c
 LOCUS
 DEFINITION
 mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AA833865 512 bp mRNA EST 07-APR-1998
 o462h07.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1372573,
 mRNA sequence.
 AA833865
 EST.
 AA833865.1 GI:2908633
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 512)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On May 8, 1995 this sequence version replaced gi:800227.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 646 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 485.

FEATURES
 source
 1. 512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1372573"
 /clone_lib="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTCAAGTGGGAGCGCGCTCATTTTTTTTTTTTTT-
 3']. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 102 a 127 c 141 g 142 t

ORIGIN

Query Match 36.1%; Score 504; DB 39; Length 512;
 Best Local Similarity 99.0%; Pred. No. 1.9e-115;
 Matches 507; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 887 cgcacaaagaatccaaagccacagctgagtcaggagagagcgttg 946
 Db 512 CGCAAAAGAAATCCAGGCCACACAGCTGAGTCTGTAGGACAGAGTGAGCGACGCTTTG 453

Qy 947 cctccgacaaagggcccttccaggacacgaagggcaagagagaccttctctaaggagact 1006
 Db 452 CTCTCAAGCAAGGCCCTTCCAGGACACGAAGGGCAAGAGAGACCTTCTTAAGAGGACT 393

Qy 1007 gcaaccacagcgccctgaggggaccacgcctccagcagagaccacagaagctcccaagtgccc 1066
 Db 392 GCAACCCAGCGCCCTGAGGGGACCAGCCTCCAGCAGGAGGCCCAAGAGCTCCCAACAGTGCC 333

Qy 1067 aagaagggagaggaagggcgacacagcagcctctggccactgcagaccccggaaggtc 1126
 Db 332 AAGAGGGGAGGAGGAAGGGGGACAGGCTACTCTGGCCACTGCAGACCCCGGAAGGTC 273

Qy 1127 aagcgtgacatcccatccttggaaccagaggggaacctcagcctcttagcagaggtctc 1186
 Db 272 AAGGCTGACATCCCATCTCTTGGAAACAGAGGGGAGCCTCAGCCCTCTAGCAGGAGCTCTC 213

Qy 1187 cttgcttgcactcaaccttttattgtcttgccctgcatctgagggtctgaattttgg 1246
 Db 212 CTGCTTGCACCTCAACCTTTCTTATTGTCTTGGCCCTGCATCTGGGGGTCTGAATTTTGG 153

Qy 1247 gagcaggaatattcgaaggcgcaaacagggccctcagcgtcttccctgcacactctcat 1306
 Db 152 GAGCAGGCAATATCTGAAGGTGCAACAGGCTCAGGCTGTTCCTCGCACACCTCTCAT 93

Qy 1307 ggttttaattgtaccccatcttccacatctttaaagctcatgtgaaaatgctgcatttt 1366
 Db 92 GGTTTTAATTGTACCCCATCTTCCACATCTTTAAAGCTCATGTGAAAATGCTGCATTTT 33

Qy 1367 taataaacatgatacatttgaaaaa 1398
 Db 32 TAATAAAGTACTACATCTGAAAAA 1

RESULT 3
 AI341866/c
 LOCUS
 DEFINITION
 mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AI341866 qt10d04.x1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1947175 3',
 mRNA sequence.
 AI341866.1 GI:4078793
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 498)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Oct 20, 1998 this sequence version replaced gi:2317572.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Sequencing by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1474 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 459.

FEATURES
 Location/Qualifiers
 1. 498
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1947175"
 /clone_lib="NCI_CGAP_GC4"
 /tissue_type="pooled germ cell tumors"


```
VERSION AA604501.1 GI:2444777
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1406948.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 879 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 471.
FEATURES
source
1. .478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1112441"
/clone_lib="NCI_CGAP_AA1"
/tissue_type="adrenal adenoma"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: adrenal gland; Vector: Bluescript SK-;
Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally.
Primer: Oligo dT. Two pooled bulk adrenal adenomas. 5'
adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor
sequence: 5' CTCGAGTCTTTTTTTTTTTT 3' Average insert
size: 1.6 kb."
BASE COUNT 100 a 125 c 134 g 119 t
ORIGIN
Query Match 33.7%; Score 471.6; DB 36; Length 478;
Best Local Similarity 99.2%; Pred. No. 2.3e-107;
Matches 474; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 908 acacagctgagctcaggacagagtgaggacgctttgctccgcaaggcccttc 967
Db 478 ACACAGCTGAGTCTCCTGAGGACAGAGTGGAGGAGCGCTTGCCTCCAGCAAGGCCCTTC 419
QY 968 aggacacgaaggcgaagagagacctctcctaagagagactgcaaccagcgccctgagggg 1027
Db 418 AGGACACGAAGGCAAGAGAGAGACTTCTTAAGAGACTGCAACCCAGCGCTCAGGGG 359
QY 1028 accagctccagcaggaccacagagtcacacagtcgcccaagaggaggaggaagggg 1087
Db 358 ACCAGGCTCCAGCAGGACCCAGAGAGTCCACAGTGCCTCCCAAGAGGGAGGAGGAGGG 299
QY 1088 cgacaggagcctctggccactcagaccgccggaaggtaaggctgacatcccatcttg 1147
Db 298 CGACAGGCGAGCTCTGGCCACTCAGACCCCGGAGAGGTCAAGGCTCAGATCCCATCTTG 239
QY 1148 gaaccagaggggacctcagcctcttagcagaggctctctcttcttgctgacatcacccttc 1207
Db 238 GAACACAGGGGACCTCAGGCTCTTAGCAGGAGGCTCTCCTTCTGCTGCATCACCCTTC 179
QY 1208 ttattgtcttgcctgcatactgggggtgctgaatttttggagcagggaataatctgaaggt 1267
Db 178 TTATTGTCTGCTGCATCTGGGGGTCTCAATTTTGGGAGCAGGCAATATCTGAAGGT 119
```

```
QY 1268 gcaaacagggccctacggctgttcctgcacaaactctcatggttttaattgtaccocatc 1327
Db 118 GCAAAAGGCGCTACGGCTGTTCCCTGCACAACCTCATGCTTAATTAATGTACCCCACT 59
QY 1328 tccacatctttaagctcatgtgaaaaatgctgaatttttaataaaactgatacatcttg 1385
Db 58 TCCACATCTTTAAAGCTCATGTGAAAATGCTGCATTTTAAATAAACTGATACATTTG 1
RESULT 6
AA831660/c 504 bp mRNA EST 18-MAR-1998
LOCUS oc83q09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1356352 3',
DEFINITION mRNA sequence.
ACCESSION AA831660
VERSION AA831660.1 GI:2904759
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 504)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2286541.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
```

```
Insert Length: 1425 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 497.
FEATURES
Location/Qualifiers
1. .504
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1356352"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

```
BASE COUNT 101 a 127 c 136 g 140 t
ORIGIN
```

```
Query Match 33.7%; Score 470.8; DB 39; Length 504;
Best Local Similarity 98.2%; Pred. No. 3.6e-107;
Matches 497; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
```



```

Db 119 TGCAAAAGAGCCCTACGGCTGTTCCTCGCACAACTCTCATGGTTTTTAATGTACCCCATC 60
QY 1327 ttccacatcttaaaagctcatgtgaaatgctgatttttaataaactgatacatcttg 1385
Db 59 TTCCACATCTTTAAAGCTCATGTGAAATAGTGTGCAATTTTAAATAAAGCTGATACATTTG 1

RESULT 10
AA280803/c 421 bp mRNA EST 14-AUG-1997
LOCUS z599a08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:711734 3',
DEFINITION mRNA sequence.
ACCESSION AA280803
VERSION AA280803.1 GI:1923483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394921.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 888 Std Error: 0.00
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 397.

FEATURES
    source
        Location/Qualifiers
            1..421
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:711734"
                /clone_lib="NCI_CGAP_GCB1"
                /tissue_type="germinal center B cell"
                /lab_host="DH10B"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand
cDNA was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 91 a 107 c 118 g 105 t
ORIGIN

Query Match 29.7%; Score 414.6; DB 31; Length 421;
Best Local Similarity 99.0%; Pred. No. 3.5e-93;
Matches 417; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 965 tcaggacagagggcgaagagacctcttaagagactgcaaccagcgccctgag 1024
Db 421 TCAGGACAGGAGGCGAAGAGAGACCTTCTTAAGAGACTGCACCCAGCGCCCTGAG 362
QY 1025 gggaccagcctccagcagaccacccagctccacacagtgcccaagaagggaggaggaag 1084
Db 361 GGGACCAAGCTCCAGCAGGAGGCCAGAGCTCCACAGTGCCTCAAGAGGGGAGGGAAG 302
QY 1085 gggcgacagggcctctggcactcagaccggcgaagtgcaagctgacatcccatcc 1144
Db 301 GGGCGACAGCGCTCTGGCCACTCGAGACCCCGGAGGCTCAAGGCTGACATCCCATCC 242

```

```

QY 1145 ttggaacagaggggacctcagcctcttagcagaggctctcttgcactcacccct 1204
Db 241 TTGGAACACAGAGGACCTCAGCCCTTTAGCAGAGGGCTCTCTTGTGCTGCACTACCCCT 182
QY 1205 ttcttattgtcttgcctgcacatctcgtggggtctgaatttttgggagcaggaatactgaa 1264
Db 181 TTCTTATTGTCTTGCCTGCATCTGGGGTCTGAATTTTGGGAGCAGGCAATATCTGAA 122
QY 1265 ggtgaaacagggccctagcgtgttcctcgcacaactcctcattggttttaattgtacccca 1324
Db 121 GGTGAAACACAGCGCTAGCGCTGTTCCTCGCACAACTCTCATGGTTTTTAATGTACCCCA 62
QY 1325 tcttcacatctttaagctcattgtgaaaaatgctgcatttttaataaactgatacatctt 1384
Db 61 TCTTCCACATCTTTAAGCTCATGTGAAAAATGCTGCATTTTAAATAAAGCTGATACATTT 2
QY 1385 g 1385
Db 1 G 1

RESULT 11
AI079757/c 412 bp mRNA EST 12-AUG-1998
LOCUS oy43b02.s1 Soares_parathyroid_tumor_MBHPA Homo sapiens cDNA clone
DEFINITION IMAGE:1668555 3', mRNA sequence.
ACCESSION AI079757
VERSION AI079757.1 GI:3416008
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 412)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Apr 18, 1995 this sequence version replaced gi:775320.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 378.
    Location/Qualifiers
        1..412
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1668555"
            /clone_lib="Soares_parathyroid_tumor_MBHPA"
            /tissue_type="parathyroid tumor"
            /dev_stage="adult"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Organ: parathyroid gland; Vector: pT7T3D
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5'-
TGTTACCAATCTGAAGTGGAGCGCGCCACCAATTTTCTTTTCTTTTCTTTT
T-3'] double-stranded cDNA was size selected, ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento

```


Qy 1368 aataaactgatacatattgaa 1387
 |||||||||||||||||||
 Db 20 NATAACTGATACATTGAA 1

RESULT 15

AI373305/c

LOCUS

DEFINITION q245a03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2029804 3',

mRNA sequence. EST 16-FEB-1999

ACCESSION AI373305

VERSION AI373305.1 GI:4153171

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 377)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Jun 15, 1998 this sequence version replaced gi:3223423.

Contact: Robert Strausberg, Ph.D.

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/brpr/image/image.html

Insert Length: 516 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 371.

Location/Qualifiers

1..377

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2029804"

/clone.lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not 1; Site_2: Eco RI;

Plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(clones 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo."

BASE COUNT 86 a 94 c 107 g 90 t

ORIGIN

Query Match 27.0%; Score 377; DB 45; Length 377;
 Best Local Similarity 100.0%; Pred. No. 8.1e-84;
 Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1010 acccagcgctgagggaccagctccagcaggaccagagctccacagtgcccaag 1069

|||||

Db 377 ACCCAGCGCCTGAGGGGACCAGCCTCCAGGAGGCCAGAGCTCCACAGTGCCCAAG 318

|||||

Qy 1070 aagggagaggaagggcgacagggcagctctggccactcagaccccgaaaggtcaag 1129

|||||

Db 317 AAGGGGAGGAGAGGGGCGACAGGACGCTGTGGCCACTGCAGACCCCGGAAGGTCAAG 258

|||||

Qy 1130 gctgacatccccctcttgaaccagaggggacctagcctcttagcaggaggtctcctt 1189

|||||

Db 257 GCTGACATCCCATCCTTGGAAACACAGAGGGGACCTCAGCCTCTTTAGCAGGAGGCTCTCCTT 198

Qy 1190 gcttgcaactcacccttcttattgtcttgccctgcatctgggggtctgaatttttggag 1249

|||||

Db 197 GCTTGCACTCACCCCTTCTTAATGTCTTGCCCTGCATCTGGGGGCTGAAATTTTGGGAG 138

|||||

Qy 1250 caggcaatatctgaaggtgcaaacagggccctacggctgttccctgcacaaactctcatgt 1309

|||||

Db 137 CAGGCAATATCTGAAGGTGCAAAACAGGCCCTACGGCTGTTCCTGCACAACTCTCATGT 78

|||||

Qy 1310 ttaattgtaccccatcttccacatctttaagctcatgtgaaaaatgctgcattttttaa 1369

|||||

Db 77 TTAAATTGTACCCCATCTTCCACATCTTTAAAGCTCATGTGAAAAAATGCTGCATTTTAA 18

|||||

Qy 1370 taaactgatacatattga 1386

|||||

Db 17 TAACTGATACATTGA 1

Search completed: May 23, 2000, 22:03:13
 Job time: 3951 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:07:12 ; Search time 6115.43 Seconds
(without alignments)
-83.672 Million cell updates/sec

Title: US-09-215-435-76
Perfect score: 526
Sequence: 1 ctgctgctgctgtgctgcac.....cttctgtcaaaaaaaaaa 526

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_ba4.*
- 5: gb_ba5.*
- 6: gb_ba6.*
- 7: gb_ba7.*
- 8: gb_ba8.*
- 9: gb_ba9.*
- 10: gb_ba10.*
- 11: gb_ba11.*
- 12: gb_ba12.*
- 13: gb_ba13.*
- 14: gb_ba14.*
- 15: gb_ba15.*
- 16: gb_ba16.*
- 17: gb_ba17.*
- 18: gb_ba18.*
- 19: gb_ba19.*
- 20: gb_ba20.*
- 21: gb_ba21.*
- 22: gb_ba22.*
- 23: gb_ba23.*
- 24: gb_ba24.*
- 25: gb_ba25.*
- 26: gb_ba26.*
- 27: gb_ba27.*
- 28: gb_ba28.*
- 29: gb_ba29.*
- 30: gb_ba30.*
- 31: gb_ba31.*
- 32: gb_ba32.*
- 33: gb_ba33.*
- 34: gb_ba34.*
- 35: gb_ba35.*
- 36: gb_ba36.*
- 37: gb_ba37.*
- 38: gb_ba38.*
- 39: gb_ba39.*
- 40: gb_ba40.*
- 41: gb_ba41.*
- 42: gb_ba42.*
- 43: gb_ba43.*
- 44: gb_ba44.*

- 45: gb_ba45.*
- 46: gb_ba46.*
- 47: gb_ba47.*
- 48: gb_ba48.*
- 49: gb_ba49.*
- 50: gb_ba50.*
- 51: gb_ba51.*
- 52: gb_ba52.*
- 53: gb_ba53.*
- 54: gb_ba54.*
- 55: gb_ba55.*
- 56: gb_ba56.*
- 57: gb_ba57.*
- 58: gb_ba58.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	202	38.4	33458	10	HSDJ60101	AL109656 Human DNA
c 2	100.4	19.1	116380	11	HS688G8	AL031671 Human DNA
c 3	52.6	10.0	97385	11	HSDJ447F3	AL050348 Human DNA
4	48.2	9.2	555	5	E12656	E12656 Synthetic D
5	48.2	9.2	1221	9	HSALMICR	X04494 Human mRNA
6	48.2	9.2	1232	5	I05217	I05217 Sequence 7
7	48.2	9.2	1239	9	HSR	X04225 Human mRNA
8	47.2	9.0	4627	10	HSALMBG3	X54818 Human gene
9	47.2	9.0	5423	9	HUMITILC08	M88249 Human inter
c 10	47.2	9.0	211423	33	ALI37850	ALI37850 Homo sapi
11	46.8	8.9	1455	5	I63559	I63559 Sequence 6
12	45.6	8.7	191	5	A31024	A31024 DNA fragmen
c 13	45.6	8.7	191	5	A31025	A31025 DNA fragmen
c 14	45.6	8.7	191180	42	AC011161	AC011161 Homo sapi
15	45.2	8.6	419	3	MEELACP	AJ000490 Macropus
c 16	45.2	8.6	153206	40	AC008356	AC008356 Homo sapi
17	44.8	8.5	938	3	S61902	S61902 tissue fact
18	44.8	8.5	1060	3	OCLIPOPR	X54708 Rabbit mRNA
19	44.6	8.5	10057	4	CHKCOLAVI	M24282 Chicken alp
20	44	8.4	435	5	E13092	E13092 DNA encodin
c 21	43.4	8.3	110000	32	PFMAL4P1_1	Continuation (2 of
22	42.2	8.0	418	5	AI6768	AI6768 variant of
23	42.2	8.0	418	5	I42457	I42457 Sequence 5
24	42.2	8.0	3572	7	PHTCYTOX	L04457 Phytophor
25	42.2	8.0	10558	9	HSCOLLVI3	X52022 H.sapiens R
26	41.4	7.9	100310	32	F17A13	AL096692 Arabidops
27	40.8	7.8	412	5	A30170	A30170 cDNA for ap
28	40.8	7.8	412	5	A30174	A30174 cDNA for ap
29	40.8	7.8	412	5	I33234	I33234 Sequence 70
30	40.8	7.8	412	5	I33236	I33236 Sequence 74
31	40.8	7.8	412	5	I40213	I40213 Sequence 70
32	40.8	7.8	412	5	I40215	I40215 Sequence 74
33	40.8	7.8	412	5	I40445	I40445 Sequence 70
34	40.8	7.8	412	5	I40447	I40447 Sequence 74
35	40.8	7.8	418	5	A30162	A30162 cDNA for ap
36	40.8	7.8	418	5	I33226	I33226 Sequence 58
37	40.8	7.8	418	5	I40205	I40205 Sequence 58
38	40.8	7.8	418	5	I40437	I40437 Sequence 58
39	40.8	7.8	1702	4	D87752	D87752 Xenopus lae
40	40.4	7.7	906	24	E10379	E10379 Rat LACI CD
41	40.4	7.7	1198	3	BTU35642	U35642 Bos taurus
42	40.4	7.7	1226	12	RATTFPI	D10926 Rattus norv
43	40.4	7.7	1228	5	E04404	E04404 cDNA sequen
44	40.2	7.6	1157	12	AB006444	AB006444 Cavia por
c 45	39.8	7.6	2377	34	DDISP70A	M26238 D.discoideu

ALIGNMENTS

QY 435 tttgtctatttctgacctagtttcttcttcttctgaaataaactgtatgatcata 494
|||||
Db 4664 TGTTCCTATTTCTGACCTAGTTTGTCTTCTCTGAAATTAACCTGATGATCA 4605
|||||
QY 495 gaatgaagaagctctttctgtca 516
|||||
Db 4604 GAATGAAAGAGTCTTTCTGTCA 4583
|||||

RESULT 2
HS688G8/c
LOCUS
DEFINITION
Human DNA sequence from clone RP4-688G8 on chromosome 20q11.2-12
Contains a gene similar to ribosomal protein S2, ESTs, STSS, GSSs
and a CpG Island, complete sequence.
ACCESSION
AL031671
VERSION
AL031671.12 GI:5522711
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 116380)
AUTHORS
Skuce, C.
TITLE
Direct Submission
JOURNAL
Submitted (30-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 3, 1999 this sequence version replaced gi:5531249.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
RP4-688G8 is
from the library RPCI-4 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP4-688G8.

FEATURES
source
1. .116380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q11.2-12"
/clone="RP4-688G8"
/clone_lib="RPCI-4"
/note="AluSg repeat: matches 1. .295 of consensus"
646. .942
1024. .1080
/note="3 copies 19 mer 84 conserved"
1892. .2057
/note="L1NA8 repeat: matches 6111. .6286 of consensus"
2066. .2371
/note="L1M4 repeat: matches 5037. .5378 of consensus"
2405. .3568
/note="L1MD repeat: matches -2. .1403 of consensus"

QY 435 tttgtctatttctgacctagtttcttcttcttctgaaataaactgtatgatcata 494
|||||
Db 4664 TGTTCCTATTTCTGACCTAGTTTGTCTTCTCTGAAATTAACCTGATGATCA 4605
|||||
QY 495 gaatgaagaagctctttctgtca 516
|||||
Db 4604 GAATGAAAGAGTCTTTCTGTCA 4583
|||||

RESULT 1
HSDJ60101/c
LOCUS
DEFINITION
Human DNA sequence from clone 60101 on chromosome 20, complete
sequence.
ACCESSION
AL109656
VERSION
AL109656.7 GI:6015550
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 33458)
AUTHORS
Laird, G.
TITLE
Direct Submission
JOURNAL
Submitted (08-OCT-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 7, 1999 this sequence version replaced gi:5918392.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
IMPORTANT: This
sequence is not the entire insert of clone 60101. It may be shorter
because we only sequence overlapping sections once, or longer
because we arrange for a small overlap between neighbouring
submissions.
The true left end of clone 688G8 is at 33359 in this sequence. The
true right end of clone dJ47F3 is at 100 in this sequence. This
sequence has been finished according to sequence map criteria as
follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
60101 is from the library RPCI-4 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.

FEATURES
source
1. .33458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone_lib="RPCI-4"
/clone="RP4-60101"
BASE COUNT 8572 a 6597 c 6799 g 11490 t
ORIGIN

Query Match 38.4%; Score 202; DB 10; Length 33458;
Best Local Similarity 100.0%; Pred. No. 2.1e-35;
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 gaggtgagagatgtgaactcatgaagtgtctgtctgacctcgaataaagacacaa 374
|||||
Db 4784 GAGGTGAGAGATGTGAACCTCATGAGATGTCTGTCTGCACCATCCGAAATAAGACACAA 4725
|||||
QY 375 gaaattcacagctgatttgaatctttgtaatttccataatgcttttaagcttcata 434
|||||
Db 4724 GAAATTCACAGCTGATTTTGAAATCTTTGTATATTTCCATATGCTTTAAGCTTCATA 4665
|||||

```
repeat_region 3571..3980
/note="LTR7 repeat: matches 1..448 of consensus"
repeat_region 3981..4356
/note="HERVH repeat: matches 7322..7713 of consensus"
repeat_region 4355..4473
/note="HERVH repeat: matches 5697..5819 of consensus"
repeat_region 4467..5011
/note="HERVH repeat: matches 4488..5031 of consensus"
repeat_region 5014..7467
/note="HERVH repeat: matches 1395..3993 of consensus"
misc_feature 7605..8006
/note="match: GSS: Em:AQ782401"
repeat_region 9746..10352
/note="L1ME3 repeat: matches 5437..6133 of consensus"
repeat_region 10353..10647
/note="AluY repeat: matches 1..295 of consensus"
repeat_region 10848..11115
/note="L1ME3 repeat: matches 4999..5437 of consensus"
repeat_region 11194..11746
/note="L1MEC repeat: matches 2332..2541 of consensus"
repeat_region 11951..12163
/note="MER20 repeat: matches 3..217 of consensus"
repeat_region 12167..12261
/note="L1M4 repeat: matches 1962..2053 of consensus"
repeat_region 12262..12568
/note="AluSg repeat: matches 1..312 of consensus"
repeat_region 12569..13254
/note="L1M4 repeat: matches 1248..1962 of consensus"
repeat_region 13262..15039
/note="L1P3 repeat: matches 607..2244 of consensus"
repeat_region 15063..15327
/note="AluSx repeat: matches 31..295 of consensus"
repeat_region 16417..16725
/note="AluSx repeat: matches 1..310 of consensus"
repeat_region 18397..18653
/note="AluSx repeat: matches 38..294 of consensus"
repeat_region 19627..19684
/note="29 copies 2 mer gt 87 conserved"
repeat_region 19983..20036
/note="27 copies 2 mer tc 90 conserved"
misc_feature complement(20022)
/note="Tandem repeat. Forced join. Gap size estimated by
BAMHI and HINDIII digest as approximately 190bp."
repeat_region 20786..20972
/note="TIGER1 repeat: matches 1..189 of consensus"
repeat_region 20983..21161
/note="L1M4 repeat: matches 2451..2630 of consensus"
repeat_region 21202..24316
/note="L1M4 repeat: matches 2981..6271 of consensus"
repeat_region 24317..24611
/note="AluY repeat: matches 1..295 of consensus"
repeat_region 24612..25973
/note="L1M4 repeat: matches 1626..2981 of consensus"
repeat_region 25974..26286
/note="AluY repeat: matches 1..309 of consensus"
repeat_region 26287..26626
/note="L1M4 repeat: matches 1268..1626 of consensus"
repeat_region 26711..26897
/note="MLT1J repeat: matches 25..207 of consensus"
repeat_region 28518..28826
/note="AluSg repeat: matches 1..310 of consensus"
repeat_region 31245..31485
/note="MIR repeat: matches 20..260 of consensus"
repeat_region 31628..31839
/note="I2 repeat: matches 2537..2749 of consensus"
repeat_region 31904..32145
/note="L1M4 repeat: matches 3952..4188 of consensus"
repeat_region 32367..32488
/note="MER3 repeat: matches 62..208 of consensus"
repeat_region 32607..32648
/note="MER3 repeat: matches 1..42 of consensus"
misc_feature complement(32649..32941)
/note="match: STS: Em:294589 Em:294589"
```

```
repeat_region 33715..33869
/note="MIR repeat: matches 11..170 of consensus"
repeat_region 34807..34940
/note="TIGER1 repeat: matches 2280..2418 of consensus"
repeat_region 35130..36430
/note="TIGER1 repeat: matches 941..2317 of consensus"
repeat_region 36433..36500
/note="TIGER2 repeat: matches 2654..2708 of consensus"
repeat_region 36501..36801
/note="AluSg repeat: matches 1..298 of consensus"
repeat_region 36802..36854
/note="TIGER2 repeat: matches 2598..2654 of consensus"
repeat_region 36853..36899
/note="FLAM repeat: matches 87..133 of consensus"
repeat_region 36901..37114
/note="AluJo repeat: matches 1..220 of consensus"
repeat_region 37115..37312
/note="TIGER2 repeat: matches 2424..2621 of consensus"
repeat_region 37320..37366
/note="TIGER2 repeat: matches 3..50 of consensus"
repeat_region 37367..38301
/note="TIGER1 repeat: matches 1..943 of consensus"
misc_feature complement(38419..38714)
/note="match: GSS: Em:AQ672639"
repeat_region 38731..39068
/note="L1MEC repeat: matches 1836..2170 of consensus"
repeat_region 39069..39376
/note="AluSx repeat: matches 1..307 of consensus"
repeat_region 39377..39414
/note="L1MEC repeat: matches 2170..2207 of consensus"
repeat_region 39645..39929
/note="AluSg repeat: matches 1..292 of consensus"
repeat_region 39965..40438
/note="L1MEC repeat: matches 2364..2486 of consensus"
repeat_region 40733..41303
/note="MER68B repeat: matches 1..569 of consensus"
repeat_region 41339..41813
/note="HAL1 repeat: matches 393..911 of consensus"
repeat_region 41906..42072
/note="L1MC4 repeat: matches 7673..7842 of consensus"
repeat_region 42684..42965
/note="AluSg repeat: matches 1..295 of consensus"
repeat_region 42967..43238
/note="L1M5A repeat: matches 6016..6291 of consensus"
repeat_region 43239..43531
/note="AluSg repeat: matches 3..294 of consensus"
repeat_region 43532..44333
/note="L1M5A repeat: matches 5202..6016 of consensus"
repeat_region 44334..44698
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 44699..45130
/note="L1M5A repeat: matches 4764..5202 of consensus"
repeat_region 45131..45447
/note="AluSg repeat: matches 1..312 of consensus"
repeat_region 45448..46910
/note="L1M5A repeat: matches 3386..4764 of consensus"
repeat_region 46911..47063
/note="MER21B repeat: matches 323..482 of consensus"
repeat_region 47105..47374
/note="L1M4 repeat: matches 2562..2839 of consensus"
repeat_region 47980..48054
/note="MER21B repeat: matches 328..420 of consensus"
repeat_region 48412..48487
/note="38 copies 2 mer ta 67 conserved"
repeat_region 51887..52069
```

```
Query Match 19.1%; Score 100.4; DB 11; Length 116380;
Best Local Similarity 76.4%; Pred. No. 5.7e-13;
Matches 136; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

Qy 325 gatgtgaactcatgaagtgtgtctgtcaccatccgaataaagacacaaagaaattcag 384
   ||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 42634 GAAAGAATAAGAGGAGTGTCTTGATGTCCCATCTGAATAAGATGCACACACATCCAG 42575
```



```

RESULT 5
HSALMICR 1221 bp mRNA PRI 30-MAR-1995
LOCUS Human mRNA for alpha-1-microglobulin and HI-30.
DEFINITION
ACCESSION X04494
VERSION X04494.1 GI:24478
KEYWORDS alpha-1-microglobulin; glycoprotein; proteinase inhibitor; serum
SOURCE protein; signal peptide.
ORGANISM human.
REFERENCE 1 (bases 1 to 1221)
AUTHORS Kaumeyer,J.F., Polazzi,J.O. and Kotick,M.P.
TITLE The mRNA for a proteinase inhibitor related to the HI-30 domain of
inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin
(protein HC)
JOURNAL Nucleic Acids Res. 14 (20), 7839-7850 (1986)
MEDLINE 87040757
COMMENT Data kindly reviewed (11-MAY-1987) by Polazzi J.O.
FEATURES
source
1..1221
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
46..1104
/notes="precursor polypeptide"
/codon_start=1
/protein_id="CAA28182.1"
/db_xref="GI:24478"
/db_xref="SWISS-PROT:P02760"
/translation="MRSGLALLLLSACLAVSAGVPVTPPDNIQVOENFISRIYKWK
YNAIGSTCPWLKIMDRVTSLVLEGATAEISMTSRWRKGVCEFTSGAYEKTD
TDGKFLYHKSKNMIMESVYVNTYDEYALFTKESRHHGPTITAKLYGRAPOLRET
LQDFRVACQGVIPDSIFTHADRGECVPGQEPPELIPRRRAVLQEEEGSGG
QLVETVTKEDSCQLGYSAGPCMGMTSRYFNGTSMACETFYQGGCMGNFNFTKE
CLQTRTVAACNLPYVRGPCRAFIQLWAFDAVKGKCVLPFYGGCGGNKIFYSEKCR
EYGVPGDGDDELLRFNS"
46..102
sig_peptide
/product="pot. signal peptide (AA -19 to -1)"
mat_peptide
103..1101
/product="alpha-1-microglobulin-HI-30 protein (AA 1-333)"
mat_peptide
103..651
/product="alpha-1-microglobulin"
mat_peptide
661..1095
/product="HI-30"
misc_feature
724..891
/notes="domain I"
misc_feature
892..1095
/notes="domain II"
misc_feature
1188..1192
/notes="put. polyA signal"
BASE COUNT 296 a 322 c 363 g 240 t
ORIGIN

Query Match 9.2%; Score 48.2; DB 9; Length 1221;
Best Local Similarity 57.7%; Pred. No. 0.29;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 134 aagatccctgcaaatggacatgaatttggaaagctgtctatgaagttcacttagatatt 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 728 AAGATTCTGCCAGCTGGGTACVCGCGCGTCCCTGCATGGGAATGACCGAGGATATT 787

Qy 194 tctacaagaacctccaagaatgtgaaactttgtctctcgcgtgtaaggcaacc 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 788 TCTATAATGGTACATCCATCGCCCTGTGAGACTTTCCAGTACGGCGGCTGCATGGGCAACG 847

Qy 254 ttaacaacttcaagcttaaaatagaacgt 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 848 GTAACAACCTTCGTACACAGAAAAGGAGTGT 876

```

```

RESULT 6
I05217 1232 bp PAT 02-DEC-1994
LOCUS Sequence 7 from Patent EP 0255011.
DEFINITION
ACCESSION I05217
VERSION I05217.1 GI:591138
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1232)
AUTHORS Kaumeyer,J.F., Kotick,M.P. and Polazzi,J.O.
TITLE Human inter-alpha-trypsin inhibitor gene
JOURNAL Patent: EP 0255011-A2 7 03-FEB-1988;
FEATURES Location/Qualifiers
source
1..1232
/organism="unknown"
BASE COUNT 300 a 323 c 365 g 244 t
ORIGIN

Query Match 9.2%; Score 48.2; DB 5; Length 1232;
Best Local Similarity 57.7%; Pred. No. 0.29;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 134 aagatccctgcaaatggacatgaatttggaaagctgtctatgaagttcacttagatatt 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 733 AAGATTCTGCCAGCTGGGTACTCGCGCGTCCCTGCATGGGAATGACCGAGGATATT 792

Qy 194 tctacaagaacctccaagaatgtgaaactttgtctctcgcgtgtaaggcaacc 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 793 TCTATAATGGTACATCCATCGCCCTGTGAGACTTTCCAGTACGGCGGCTGCATGGGCAACG 852

Qy 254 ttaacaacttcaagcttaaaatagaacgt 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 853 GTAACAACCTTCGTACACAGAAAAGGAGTGT 881

RESULT 7
HSHCR 1239 bp mRNA PRI 11-AUG-1992
LOCUS Human mRNA for protein HC (alpha-1-microglobulin).
DEFINITION
ACCESSION X04225
VERSION X04225.1 GI:32046
KEYWORDS alpha-1-microglobulin; glycoprotein; microglobulin; protein HC;
signal peptide.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1231)
AUTHORS Traboni,C. and Cortese,R.
TITLE Sequence of a full length cDNA coding for human protein HC (alpha 1
microglobulin)
JOURNAL Nucleic Acids Res. 14 (15), 6340 (1986)
MEDLINE 86312901
REMARK revised by [2]
REFERENCE 2 (bases 1 to 1239)
AUTHORS Cortese,R.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1992) Cortese R., IRBM, Via Pontina km 30.600,
Pomezia (Rome), Italy
FEATURES Location/Qualifiers
source
1..1239
/organism="Homo sapiens"
/db_xref="taxon:9606"
73..1131
/codon_start=1
/product="HC polypeptide"
/protein_id="CAA27803.1"
/db_xref="GI:32047"
/db_xref="SWISS-PROT:P02760"
/translation="MRSGLALLLLSACLAVSAGVPVTPPDNIQVOENFISRIYKWK"
CDS

```

YNLAIGSTCPWLKIMDRMTVSTVLIGEGATEAEISMTSTRWRKGVCEETSGAYEKTD
 TQCKFLYHKSQVNIWESVYVHTNYDEYAIFLTKEFSRHGPTITAKLYGRAPQLRET
 LLDQFVAGVGLPDSIFTMADRGECVPGQPEPILIPVRRVAVLPQEEGSGGG
 QLVTEYTKEDSCOLGYSAGPCMGWTSRYFYNGTSMACETFYGGCMGNPNFYTEKE
 CLQCTRTVAACNLPVIRGCPRAFQIQLWAFDAVKGKCVLPFYGGCQGNKFNFYSEKCR
 EYCGVPGDGBELLRFN"

sig_peptide

/product="HC polypeptide"

mat_peptide

/product="HC polypeptide"

polyA_site

BASE COUNT 290 a 328 c 375 g 246 t

ORIGIN

Query Match 9.2%; Score 48.2; DB 9; Length 1239;
 Best Local Similarity 57.7%; Pred. Mismatches 63; Indels 0; Gaps 0;
 Matches 86; Conservative 0;

QY 134 aagatccctgcaaatggacatgaattttggaagctgtatgaagttcactttagatatt 193

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 755 AAGATTCCTGCAGCTGGGTACTCGCGCGTCCCTGCATGGGATGACCAGCAGGTATT 814

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 194 tctacaacagaaacctccaaagatgtgaaacttttgccttcctcggtgttaatggcaacc 253

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 815 TCTATAATGGTACATCCATGCCCTGTGAGACTTTCAGTACGGGGCTGCATGGCAACG 874

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 254 ttaacaacttcaagcttaataatgaacgt 282

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 875 GTAACAACCTTCGTACACAGAAAGGAGTGT 903

RESULT 8

HSAIMBG3

LOCUS

DEFINITION HSAIMBG3 4627 bp DNA PRI 06-OCT-1999
 human gene for alpha-1-microglobulin-bikunin, exons 7-10 (encoding
 linker peptide and bikunin).

ACCESSION

X54818

KEYWORDS

alpha-1-microglobulin; bikunin; glycoprotein; proteinase inhibitor.

SOURCE

HUMAN

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 4627)

Vetr.H. and Gebhard,W.

TITLE Structure of the human alpha 1-microglobulin-bikunin gene

JOURNAL Biol. Chem. Hoppe-Seyler 371 (12), 1185-1196 (1990)

MEDLINE 91214554

2 (bases 1 to 4627)

Gebhard,W.

Direct Submission

TITLE Submitted (09-OCT-1990) Gebhard W., Universitaet Muenchen, Klinikum

JOURNAL Grosshadern, HNO Forschungslabor, Marchioninistr 15, 8000 Muenchen

70, Germany

Sequence is approximately 6 kbp downstream of x54817.

COMMENT

FEATURES

Location/Qualifiers

source

1..4627

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="leukocyte"

/clone_lib="EMBL3 vector"

prim_transcript

<1..2801

/gene="alpha-1-microglobulin-bikunin"

<1..171

/gene="alpha-1-microglobulin-bikunin"

/number=6

1..2801

/gene="alpha-1-microglobulin-bikunin"

172..253

/gene="alpha-1-microglobulin-bikunin"

/number=7

/product="linker peptide and bikunin"

mRNA

join(172..253,1339..1506,1832..2005,2658..2801)

/partial

/gene="alpha-1-microglobulin-bikunin"

/usedin=X54816:ALMB_mRNA_a

/usedin=X54816:ALMB_mRNA_b

join(172..253,1339..1506,1832..2005,2658..2689)

/partial

/gene="alpha-1-microglobulin-bikunin"

/note="alpha-1-microglobulin"

/codon_start=1

/usedin=X54816:ALMB_cds

/product="bikunin"

/protein_id="CAA38587.1"

/db_xref="GI:579676"

/db_xref="SPTREMBL:P78491"

/translation="RYRRVAVLPQEEGGGGQLVTEYTKEDSCOLGYSAGPCMGWTS

RYFYNGTSMACETFYGGCMGNPNFYTEKECLQCTRTVAACNLPVIRGCPRAFQIQLW

AFDAVKGKCVLPFYGGCQGNKFNFYSEKREYCGVPGDGBELLRFN"

intron

254..1338

/gene="alpha-1-microglobulin-bikunin"

/number=7

1339..1506

/gene="alpha-1-microglobulin-bikunin"

/number=8

/product="bikunin"

1507..1831

/gene="alpha-1-microglobulin-bikunin"

/number=8

1832..2005

/gene="alpha-1-microglobulin-bikunin"

/number=9

/product="bikunin"

2006..2657

/gene="alpha-1-microglobulin-bikunin"

/number=9

2658..2801

/gene="alpha-1-microglobulin-bikunin"

/number=10

/product="bikunin"

2773..2778

/gene="alpha-1-microglobulin-bikunin"

/repeat_unit

/rpt_family="Alu"

BASE COUNT 1088 a 1133 c 1292 g 1108 t 6 others

ORIGIN

Query Match 9.0%; Score 47.2; DB 10; Length 4627;

Best Local Similarity 57.4%; Pred. No. 0.43;

Matches 85; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 135 agatccctgcaaatggacatgaattttggaagctgtatgaagttcactttagatatt 194

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1337 AGATCTGCGCAGCTGGGCTACTCGCGCGTCCCTGCATGGGAATGACCAGCAGGTATT 1396

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 195 ctacaacagaacctccaaaagatgtgaaacttttcttccggctgtatggcaacct 254

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1397 CTATAATGGTACATCCATGGCGCTGTGAGACTTTCACGCGCGCTGCATGGCAACGG 1456

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 255 taacaacttcaagcttaataatagaacgt 282

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1457 TAACAACCTTCGTACAGAAAGGAGTGT 1484

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9

HUMITILC08

LOCUS HUMITILC08 5423 bp DNA PRI 06-JAN-1995

DEFINITION Human inter-alpha-trypsin inhibitor light chain (ITI) gene, exons

7-10.

ACCESSION M88249 M27417

VERSION M88249.1 GI:186598

Contig_ID: 00517 Length: 13876bp
Contig_ID: 00590 Length: 4526bp
Contig_ID: 00618 Length: 11755bp
Contig_ID: 00682 Length: 8988bp
Contig_ID: 00690 Length: 5376bp
Contig_ID: 00737 Length: 3655bp
Contig_ID: 00824 Length: 1275bp
Contig_ID: 00896 Length: 1147bp
Contig_ID: 00903 Length: 9450bp
Contig_ID: 00947 Length: 11063bp
Contig_ID: 00976 Length: 2843bp
Contig_ID: 00981 Length: 15080bp
Contig_ID: 00995 Length: 5180bp
Contig_ID: 01026 Length: 8240bp
Contig_ID: 01032 Length: 11217bp
Contig_ID: 01035 Length: 1076bp
Contig_ID: 01101 Length: 3875bp
Contig_ID: 01103 Length: 6085bp
Contig_ID: 01122 Length: 1150bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 8133: contig of 8133 bp in length
* 8134 8933: gap of 800 bp
* 8934 11280: contig of 2347 bp in length
* 11281 12080: gap of 800 bp
* 12081 18554: contig of 6474 bp in length
* 18555 19354: gap of 800 bp
* 19355 32024: contig of 12670 bp in length
* 32025 32824: gap of 800 bp
* 32825 34894: contig of 2070 bp in length
* 34895 35694: gap of 800 bp
* 35695 38630: contig of 2936 bp in length
* 38631 39430: gap of 800 bp
* 39431 41674: contig of 2244 bp in length
* 41675 42474: gap of 800 bp
* 42475 45255: contig of 2781 bp in length
* 45256 46055: gap of 800 bp
* 46056 49866: contig of 3811 bp in length
* 49867 50666: gap of 800 bp
* 50667 55149: contig of 4483 bp in length
* 55150 55949: gap of 800 bp
* 55950 57660: contig of 1711 bp in length
* 57661 58460: gap of 800 bp
* 58461 68129: contig of 9669 bp in length
* 68130 68929: gap of 800 bp
* 68930 72486: contig of 3557 bp in length
* 72487 73286: gap of 800 bp
* 73287 87162: contig of 13876 bp in length
* 87163 87962: gap of 800 bp
* 87963 92488: contig of 4526 bp in length
* 92489 93288: gap of 800 bp
* 93289 105043: contig of 11755 bp in length
* 105044 105843: gap of 800 bp
* 105844 114741: contig of 8898 bp in length
* 114742 115541: gap of 800 bp
* 115542 120917: contig of 5376 bp in length
* 120918 121717: gap of 800 bp
* 121718 125372: contig of 3655 bp in length
* 125373 126172: gap of 800 bp
* 126173 127447: contig of 1275 bp in length
* 127448 128247: gap of 800 bp
* 128248 129394: contig of 1147 bp in length
* 129395 130194: gap of 800 bp
* 130195 139644: contig of 9450 bp in length
* 139645 140444: gap of 800 bp
* 140445 151507: contig of 11063 bp in length
* 151508 152307: gap of 800 bp

* 152308 155150: contig of 2843 bp in length
* 155151 155950: gap of 800 bp
* 155951 169030: contig of 13080 bp in length
* 169031 169830: gap of 800 bp
* 169831 174980: contig of 5150 bp in length
* 174981 175780: gap of 800 bp
* 175781 184020: contig of 8240 bp in length
* 184021 184820: gap of 800 bp
* 184821 196037: contig of 11217 bp in length
* 196038 196837: gap of 800 bp
* 196838 197913: contig of 1076 bp in length
* 197914 198713: gap of 800 bp
* 198714 202588: contig of 3875 bp in length
* 202589 203388: gap of 800 bp
* 203389 209469: contig of 6081 bp in length
* 209470 210273: gap of 804 bp
* 210274 211423: contig of 1150 bp in length.
FEATURES
source
Location/Qualifiers
1. .211423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="q31.3-33.3"
/clone_lib="RPCI-11.1"
BASE COUNT 47172 a 45728 c 44881 g 48805 t 24837 others
ORIGIN
Query Match 9.0%; Score 47.2; DB 33; Length 211423;
Best Local Similarity 57.4%; Pred. No. 0.31;
Matches 85; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 135 agatccctgcaattggacatgaattttggaagctgctatgaagttcatttagatatt 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162592 AGATTCTGCGCAGCTGCGGCTACTCGCGCGCTCCCTGCGATGGGAATGACACGAGGTATTT 162533
QY 195 ctacaacagaaacctccaaaagatgtgaaactttgtctctccgctgtaatggcaacct 254
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162532 CTATATGGGTACATTCATGCGCTGTGAGACTTCCAGTACGCGGCTGCATGGGCAACGG 162473
QY 255 taacaacttcaagcttaaaatagaacgt 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162472 TAACAACCTTCGTACAGAAAGAGTGT 162445
RESULT 11
LOCUS I63559 1455 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 6 from patent US 5663143.
ACCESSION I63559
VERSION I63559.1 GI:2481132
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1455)
AUTHORS Ley,A.Charles, Ladner,R.Charles, Guterman,S.Kosow,
Roberts,B.Lindsay, Markland,W. and Kent,R.Baribault.
TITLE Engineered human-derived kunitz domains that inhibit human
neutrophil elastase
JOURNAL Patent: US 5663143-A 6 02-SEP-1997;
FEATURES
source
Location/Qualifiers
1. .1455
/organism="unknown"
BASE COUNT 337 a 316 c 366 g 436 t
ORIGIN
Query Match 8.9%; Score 46.8; DB 5; Length 1455;
Best Local Similarity 55.6%; Pred. No. 0.59;
Matches 90; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:59:37 ; Search time 282.38 seconds
(without alignments)
466.042 Million cell updates/sec

Title: US-09-215-435-76
Perfect score: 526
Sequence: 1 ctgctgctgctgtgctgcac.....ctttgtcacaataaaaaa 526

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	256.4	48.7	271	X51569	Human secreted pro
2	48.2	9.2	222	T15870	Encodes human urin
3	48.2	9.2	441	V40046	rUTI cDNA SEQ ID N
4	48.2	9.2	1232	N81432	Sequence encoding
5	48.2	9.2	1255	T78950	Anti-trypsin inhib
6	45.8	8.9	285	T35164	ITi-Di Kunitz doma
7	45.6	8.7	191	Q66878	Encodes human bku
8	45.6	8.7	372	T79088	Ep7-d21-RPDF-52-55
9	45.6	8.7	467	T79081	Insert from plasmid
10	44	8.4	165	T79079	RPDF-Kunitz domain
11	44	8.4	372	T79080	Epl-d21-RPDF-52-55
12	44	8.4	435	T79083	Epl-UTI-RPDF-52-55
13	44	8.4	467	T79086	Insert from plasmid
14	44	8.4	530	T79087	Insert from plasmid
15	42.6	8.1	186	Q21070	Synthetic aprotinin
16	42.2	8.0	201	Q53999	Sequence encoding
17	42.2	8.0	210	Q53998	Sequence encoding
18	42.2	8.0	418	Q46652	pRFA-1745 412 bp E
19	41	7.8	153	T79077	Anti-trypsin inhib
20	40.8	7.8	412	Q24424	Encodes aprotinin
21	40.8	7.8	412	Q24426	Encodes aprotinin
22	40.8	7.8	418	Q24420	Encodes aprotinin
23	40.4	7.7	906	T10945	Lipoprotein-associ
24	40.4	7.7	1228	Q39947	LACI gene. Rat pro
25	39.2	7.5	507	Q24425	Encodes aprotinin
26	38.6	7.3	418	Q46936	CDNA encoding synt
27	37.6	7.1	418	Q05920	Encodes plac212spx
28	37.6	7.1	508	Q24427	Encodes aprotinin
29	37.6	7.1	508	Q24428	Encodes aprotinin
30	37	7.0	418	Q46935	CDNA encoding synt
31	37	7.0	110000	V21209_09	Continuation (10 o
32	36.6	7.0	1642	Q71213	Human type II GnrH
33	36.2	6.9	950	Q51001	LACI gene. Gene fr
34	36	6.8	177	N91756	Aprotinin analogue

35 36 6.8 177 1 Q05923 Synthetic aprotini
36 36 6.8 180 1 N70838 Sequence from expr
37 36 6.8 181 1 N90673 Synthetic gene enc
38 36 6.8 181 1 N90674 Synthetic gene enc
39 36 6.8 182 1 N90675 Synthetic gene enc
40 36 6.8 182 1 N90676 Synthetic gene enc
41 36 6.8 191 1 Q05918 Synthetic aprotini
42 36 6.8 207 1 N70841 Sequence of synth
43 36 6.8 412 1 Q05919 Encodes plac212spx
44 36 6.8 418 1 Q05921 Encodes plac212spx
45 36 6.8 418 1 Q05922 plac212spx3 leader

ALIGNMENTS

RESULT 1

X51569
ID X51569 standard; cDNA; 271 BP.
AC X51569;
DE 21-JUN-1999 (first entry)
DE Human secreted protein 5', EST SEQ ID NO:148.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906549-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1231.
PR 01-AUG-1997; US-905279.
PA (GSEF) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153779/13.
DR P-PSDB; Y12791.
PT New nucleic acids encoding human secreted proteins - obtained from
PT CDNA libraries derived from testis, ovary, uterus and spleen tissue
PS Claim 1; Page 253-254; 522pp; English.
CC X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12681 to Y12913,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 271 BP; 75 A; 57 C; 55 G; 84 T;

Query Match 48.7%; Score 256.4; DB 1; Length 271;
Best Local Similarity 99.6%; Pred. No. 3.6e-60;
Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ctgctgctgctgtgctgcaccatgaagcttcgcaagctgggtttcttcttaagattcttc 60
|||||
Db 11 CTGCCCTGCTGCTGCTGCACCAAGTCTGCCAAGCTGGGATTTCTTCTTAAGATTCTTC 70
QY 61 atctctgtctcattgaataccctgttattgggtggttaataaaaaattcgaggagagata 120
|||||
Db 71 ATCTTCTGCTCATTAATAACCCCTGTTATTGGGTGGTGAATAAATAATTCGGGAGAGATA 130
QY 121 tctggagacctcaaaagatccctgcacatggacatgaattttggagctgctatgaagt 180


```

other respiratory disorders
Example 1; Page 43; 105pp; English.
Genetically engineered human derived Kunitz domains can be used to
inhibit human neutrophil elastase, an enzyme involved in the
elimination of pathogens and the restructuring of connective tissue.
In cases of reduction of the circulating alpha-1-protease inhibitor
(API or alpha-1antitrypsin), or the inactivation of API by oxidation
(smothers emphysema), extensive destruction of the lung tissue may
result from uncontrolled elastolytic activity of human neutrophil
elastase. Other respiratory disorders such as cystic fibrosis are
thought to be caused by human neutrophil elastase release by
neutrophils. The genetically engineered human derived Kunitz
domains can be used to treat such respiratory disorders. See
R99146-R99211. Fusion genes were used in the production of the
Kunitz domain derivatives. This sequence is a fragment of a fusion
gene that encodes the M13 III signal peptide, IPI-D1 Kunitz domain
and the mature M13 III protein. Substitutions in the IPI-D1 domain
can be made by use of BglI, EagI, NcoI, StyI, PstI and Kasi
restriction sites found within the sequence or by
oligonucleotide-directed mutagenesis of single-stranded phage DNA.
The resulting peptide derivatives can be expressed as intact human
neutrophil elastase binding proteins or can be incorporated into
larger proteins as a domain.
Sequence 285 BP; 74 A; 67 C; 68 G; 76 T;

Query Match 8.9%; Score 46.8; DB 1; Length 285;
Best Local Similarity 55.6%; Pred. No. 0.00068;
Matches 90; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 121 tgtgagacacccaagcctcgaattgacatgaattttggaagctgtatgaagt 180
DB 46 TCTGGCGCCAAAGAAAGAACTCTTGCCAGCTGGGTACTGCGCGCGTCCCTGCATGGGAATG 105

QY 181 cactttagatattctcaacagacaccccaagaatgtgaaactttgtctctccgcg 240
DB 106 ACCAGCAGGATTTCTATATGGTACATCCATGCGCTGTGAGACTTCCAGTACGGCGGC 165

QY 241 tgaatggcaaccttaacacactcaagctaaatagaacgt 282
DB 166 TGCATGGGCAACGGTAACTTCGTCACAGAAAGGAGTGT 207

RESULT 7
Q06878
ID Q06878 standard; DNA; 191 BP.
DE 05-MAR-1991 (first entry)
DE Encodes human bikunin domain 1-derived protease inhibitor.
KW bikunin; serine protease inhibitor; rheumatoid arthritis; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT mutation 57..59
FT /tag= a
FT /note= "substitute with codon for Leu, Ile, Val,
FT Arg, Phe, Tyr, Thr or Lys"
FT mutation 63..65
FT /tag= b
FT /note= "substitute with codon for Leu, Arg, Ile,
FT Val or Lys"
FT EP-401508-A.
FT 12-DEC-1990.
FT 01-MAY-1990; 108284.
FT 13-MAY-1989; DE-915689.
FT 18-JAN-1990; DE-001244.
FT (FARB ) BAYER AG.
FT Fritz H, Gebhard W, Das R;
FT WPI: 90-369615/50.
FT P-PSDB; R08291.
FT Protease-inhibitor variant derived from human bikunin - prodn.
FT used as therapeutic for treatment of eg. rheumatoid arthritis
PS Disclosure; Fig 2; 29pp; German.
CC The 5' end overhangs the 3' end of the complementary strand by 5

other respiratory disorders
bases: the 5' end of the complementary strand overhangs the 3' end
of the sense strand by AGCTG. The sequence encodes residues 19-77 of
mature bikunin in which Met(36) and/or Met(38) are substituted by
another naturally occurring amino acid. The sequence comprises domain
1 of bikunin and is synthesised from two synthetic oligonucleotides.
The substitutions are achieved by site-directed mutagenesis.
Plasmids containing the sequences are used to transform bacteria
and the protease inhibitory fragments are secreted as fusion
proteins. The fragments can be used for the treatment of emphysema,
septic shock, rheumatoid arthritis and coagulation disorders.
See also Q06879.
Sequence 191 BP; 49 A; 43 C; 46 G; 53 T;

Query Match 8.7%; Score 45.6; DB 1; Length 191;
Best Local Similarity 56.8%; Pred. No. 0.0013;
Matches 84; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 134 aagatccctcgcaattggacatgaattttggaagctgtatgaagttcactttagatt 193
DB 19 AGACTCTTGCCAGCTGGGCTACTCTGCTGGTCCGTGCATGGGTATGACTTCTCGTTACT 78

QY 194 tctacaagaacacctccaaaagaatgtgaaactttgtctctccgcgctgaatggcaacc 253
DB 79 TCTACAACGGTACTCTTATGGCTTGGCAAACTTCCAGTACGGTGGTTCATGGGTACG 138

QY 254 ttaacaacttcaagcttaaaatagaacg 281
DB 139 GTAACAACCTCTGTTACTGAAAAGAAATG 166

RESULT 8
T79081
ID T79081 standard; DNA; 372 BP.
AC T79081; 1997 (first entry)
DE Ep7-d21-RPDF-52-55 coding sequence.
KW Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW neutrophil; disease; modification; site directed mutagenesis; ss.
OS Synthetic.
PN J09124700-A.
PD 13-MAY-1997.
PF 07-NOV-1995; 288527.
PR 07-NOV-1995; JP-288527.
PA (GREC ) GREEN CROSS CORP.
DR WPI: 97-316576/29.
DR P-PSDB; W25933.
PT New protease inhibitor - useful for treating diseases involving
PT elastase
PS Disclosure; Page -; 37pp; Japanese.
CC This is the nucleotide sequence encoding a construct comprising the
CC anti-trypsin inhibitor UTI Kunitz domains 1 and 2 (see T78950) linked at
CC the 5' terminus to a sequence encoding the peptide Arg-Pro-Asp-Phe.
CC The construct lacks the sequence encoding amino acids 1-21 and the
CC Arg-Pro-Asp-Phe sequence replaces the native amino acids 22-26. The
CC construct also has the sequence encoding the active site and the
CC adjacent amino acids (amino acids 36-40 of the wild type sequence) and
CC amino acids 52 and 55 substituted. The modified protease inhibitors are
CC targeted to the protease elastase, especially from neutrophils and can
CC be used to treat diseases associated with elastase. Modifications of the
CC active site were done by site directed mutagenesis.
CC Note: this sequence is not given in the specification but is constructed
CC from the related construct Epl-d21-RPDF-52-55 (T79080).
CC Sequence 372 BP; 74 A; 98 C; 111 G; 89 T;

Query Match 8.7%; Score 45.6; DB 1; Length 372;
Best Local Similarity 56.8%; Pred. No. 0.0016;
Matches 84; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 135 agatccctcgcaattggacatgaattttggaagctgtatgaagttcactttagatt 194
DB 6 AGATTTCTGTCTCAACTGGGCTACTCGCGCGGTCCCTCGGTGCTATGTTCTTAGTATT 65
```


ID T79079 standard; DNA; 165 BP.
 AC T79079;
 DT 11-NOV-1997 (first entry)
 DE RPDI-Kunitz domain 11 coding sequence.
 KW Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
 KW neurophil; disease; modification; site directed mutagenesis; ss.
 T79079

109 G; 89 T;

```
Query Match      8.4%; Score 44; DB 1; Length 372;
Best Local Similarity 56.1%; Pred. No. 0.0042;
Matches 83; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 135 agatccctgcaaatggacatgaattttgaaagctgctatgaagttcaccttttagatattt 194
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 6 AGATTCTGTCACTGGGCTACTCGCGCGTCCCTGTCATCGCTTCTTCTAGGTATTT 65

QY 195 ctacaacagaaacctcccaaaagatgtgaaacttttcttcctcggtgttaatggcaacct 254
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 66 CTATAATGGTACATCCATCGCTCTCAGACTTTCGTACGGCGGATGCATGGCAACGG 125

QY 255 taacaacttcaagcttaaaatagaacgt 282
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 126 TAACAACCTTCGTACAGAAAAGGAGTGT 153

RESULT 12
T79083
ID T79083 standard; DNA; 435 BP.
AC T79083;
DT 11-NOV-1997 (first entry)
DE Epi-UTI-RPDF-52-55 coding sequence.
KW Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW neutrophil; disease; modification; site directed mutagenesis; ss.
OS Synthetic.
PN J09124700-A.
PD 13-MAY-1997.
PF 07-NOV-1995; 288527.
PR 07-NOV-1995; JP-288527.
PA (GREC ) GREEN CROSS CORP.
DR WPI; 97-316576/29.
PT P-PSDB; W25935.
PT New protease inhibitor - useful for treating diseases involving
PT elastase
PS Disclosure; Page 23-24; 37pp; Japanese.
CC This is the nucleotide sequence encoding a protease inhibitor construct
CC based on the anti-trypsin inhibitor URI (W25928). The novel inhibitor
CC contains the Kunitz domains 1 and 2 of URI with several amino acid
CC substitutions: amino acids 22-25 of the wild type protein (KEDS) are
CC replaced by the sequence RPDF; the active site sequence and adjacent
CC amino acids (amino acids 36-40 of the wild type sequence) are replaced
CC by the sequence IAFP; and amino acids 52 and 55 are also substituted.
CC The modified protease inhibitors are targeted to the protease elastase,
CC especially from neutrophils and can be used to treat diseases associated
CC with elastase. Modifications of the active site were done by site
CC directed mutagenesis.
SQ Sequence 435 BP; 93 A; 115 C; 130 G; 97 T;

Query Match      8.4%; Score 44; DB 1; Length 435;
Best Local Similarity 56.1%; Pred. No. 0.0044;
Matches 83; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 135 agatccctgcaaatggacatgaattttgaaagctgctatgaagttcaccttttagatattt 194
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69 AGATTCTGTCACTGGGCTACTCGCGCGTCCCTGTCATCGCTTCTTCTAGGTATTT 128

QY 195 ctacaacagaaacctcccaaaagatgtgaaacttttcttcctcggtgttaatggcaacct 254
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 129 CTATATGGTACATCCATCGCTCTCAGACTTTCGTACGGCGGATGCATGGCAACGG 188

QY 255 taacaacttcaagcttaaaatagaacgt 282
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 189 TAACAACCTTCGTACAGAAAAGGAGTGT 216

RESULT 13
T79086
ID T79086 standard; DNA; 467 BP.
AC T79086;
DT 12-NOV-1997 (first entry)
DE Insert from plasmid PHH337.
KW Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW neutrophil; disease; modification; site directed mutagenesis; ss.
OS Synthetic.
PN J09124700-A.
PD 13-MAY-1997.
PF 07-NOV-1995; 288527.
PR 07-NOV-1995; JP-288527.
PA (GREC ) GREEN CROSS CORP.
DR WPI; 97-316576/29.
PT P-PSDB; W25937.

Query Match      8.4%; Score 44; DB 1; Length 467;
Best Local Similarity 56.1%; Pred. No. 0.0045;
Matches 83; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 135 agatccctgcaaatggacatgaattttgaaagctgctatgaagttcaccttttagatattt 194
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 72 AGATTCTGTCACTGGGCTACTCGCGCGTCCCTGTCATCGCTTCTTCTAGGTATTT 131

QY 195 ctacaacagaaacctcccaaaagatgtgaaacttttcttcctcggtgttaatggcaacct 254
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 132 CTATAATGGTACATCCATCGCTCTCAGACTTTCGTACGGCGGATGCATGGCAACGG 191

QY 255 taacaacttcaagcttaaaatagaacgt 282
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 192 TAACAACCTTCGTACAGAAAAGGAGTGT 219

RESULT 14
T79087
ID T79087 standard; DNA; 530 BP.
AC T79087;
DT 12-NOV-1997 (first entry)
DE Insert from plasmid PHH355.
KW Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW neutrophil; disease; modification; site directed mutagenesis; ss.
OS Synthetic.
PN J09124700-A.
PD 13-MAY-1997.
PF 07-NOV-1995; 288527.
PR 07-NOV-1995; JP-288527.
PA (GREC ) GREEN CROSS CORP.
DR WPI; 97-316576/29.
PT P-PSDB; W25937.

Query Match      8.4%; Score 44; DB 1; Length 467;
Best Local Similarity 56.1%; Pred. No. 0.0045;
Matches 83; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 135 agatccctgcaaatggacatgaattttgaaagctgctatgaagttcaccttttagatattt 194
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 72 AGATTCTGTCACTGGGCTACTCGCGCGTCCCTGTCATCGCTTCTTCTAGGTATTT 131

QY 195 ctacaacagaaacctcccaaaagatgtgaaacttttcttcctcggtgttaatggcaacct 254
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 132 CTATAATGGTACATCCATCGCTCTCAGACTTTCGTACGGCGGATGCATGGCAACGG 191

QY 255 taacaacttcaagcttaaaatagaacgt 282
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 192 TAACAACCTTCGTACAGAAAAGGAGTGT 219
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:54:26 ; Search time 203.6 seconds
(without alignments)
335.816 Million cell updates/sec

Title: US-09-215-435-76

Perfect score: 526

Sequence: 1 ctgcctgctgctgtgcac.....ctttctgcaaaaaaaaaa 526

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46.8	8.9	1455	1	US-08-358-160-6
2	42.6	8.1	186	6	PCT-US91-04744-1
3	42.6	8.1	186	6	PCT-US91-04744-3
4	42.6	8.1	186	7	5258302-1
5	42.2	8.0	418	1	US-08-334-773A-5
6	40.8	7.8	412	1	US-08-084-718-70
7	40.8	7.8	412	1	US-08-084-718-74
8	40.8	7.8	412	1	US-08-443-976-70
9	40.8	7.8	412	1	US-08-443-976-74
10	40.8	7.8	412	1	US-08-443-977-70
11	40.8	7.8	412	1	US-08-443-977-74
12	40.8	7.8	418	1	US-08-084-718-58
13	40.8	7.8	418	1	US-08-443-976-58
14	40.8	7.8	418	1	US-08-443-977-58
15	39.2	7.5	508	1	US-08-084-718-72
16	39.2	7.5	508	1	US-08-443-976-72
17	39.2	7.5	508	1	US-08-443-977-72
18	38.6	7.3	418	1	US-08-321-658B-17
19	37.6	7.1	508	1	US-08-084-718-76
20	37.6	7.1	508	1	US-08-084-718-78
21	37.6	7.1	508	1	US-08-443-976-76
22	37.6	7.1	508	1	US-08-443-976-78
23	37.6	7.1	508	1	US-08-443-977-76
24	37.6	7.1	508	1	US-08-443-977-78
25	37	7.0	418	1	US-08-321-658B-15
26	36	6.8	176	1	US-07-598-737C-34
27	36	6.8	177	1	US-07-598-737C-32

28	36	6.8	177	1	US-08-084-718-28
29	36	6.8	177	1	US-08-084-718-33
30	36	6.8	177	1	US-08-443-976-28
31	36	6.8	177	1	US-08-443-976-33
32	36	6.8	177	1	US-08-443-977-28
33	36	6.8	177	1	US-08-443-977-33
34	36	6.8	178	1	US-08-084-718-38
35	36	6.8	178	1	US-08-084-718-40
36	36	6.8	178	1	US-08-443-976-38
37	36	6.8	178	1	US-08-443-976-40
38	36	6.8	178	1	US-08-443-977-38
39	36	6.8	178	1	US-08-443-977-40
40	36	6.8	450	2	US-08-446-646-4
41	36	6.8	470	2	US-08-446-646-6
42	36	6.8	508	1	US-08-084-718-80
43	36	6.8	508	1	US-08-084-718-82
44	36	6.8	508	1	US-08-443-976-80
45	36	6.8	508	1	US-08-443-976-82

ALIGNMENTS

RESULT 1

US-08-358-160-6

; Sequence 6, Application US/08358160

; Patent No. 5663143

; GENERAL INFORMATION:

; APPLICANT: LEY, Arthur C.

; APPLICANT: LADNER, Robert C.

; APPLICANT: GUTERMAN, Sonia K.

; APPLICANT: ROBERTS, Bruce L.

; APPLICANT: MARKLAND, William

; APPLICANT: KENT, Rachel B.

; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ

; NUMBER OF SEQUENCES: 234

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W. Suite 300

; CITY: Washington

; STATE: District of Columbia

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/358,160

; FILING DATE: 16-DEC-1994

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/133,031

; FILING DATE: 13-OCT-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/009,319

; FILING DATE: 26-JAN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/664,989

; FILING DATE: 01-MAR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/487,063

; FILING DATE: 02-MAR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/240,160

; FILING DATE: 02-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Cooper, Iver P.

; REGISTRATION NUMBER: 28,005

; REFERENCE/DOCKET NUMBER: LEV=1

; TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: synthetic DNA fragment
; US-08-358-160-6

Query Match      8.9%; Score 46.8; DB 1; Length 1455;
Best Local Similarity 55.8%; Pred. No. 0.00042;
Matches 90; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 121 tgtggagacctcaaaagatccctgcaaatggacatgaattttggaagctgctatgaagtt 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 46 TCTGGCCCAAGAGAGACTCTTGCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATG 105

QY 181 cacttttagatattctacacagaacctccaaagaatgtgaaacttttcttctccggc 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106 ACCAGCAGGTATTTCTATAATGTGTACATCCATCGCCTGTGAGACTTTCCAGTACGGCGC 165

QY 241 tgaatagcaaccttaacaacttcaagcttaaaatagaacgt 282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166 TGCATGGCAACGGTAACAACCTTCGTCACAGAAAGGAGTGT 207

RESULT 2
PCT-US91-04744-1
; Sequence 1, Application PC/TUS9104744
; GENERAL INFORMATION:
; APPLICANT: Vedvick, Thomas S
; APPLICANT: Engel, Michael E
; APPLICANT: Urcan, Mary S
; APPLICANT: Buckholz, Richard G
; APPLICANT: Kinney, Jennifer A
; TITLE OF INVENTION: Production of Aprotinin in
; TITLE OF INVENTION: Methylotrophic Yeast Cells
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 S. LaSalle St.
; CITY: Chicago
; STATE: Ill.
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/04744
; FILING DATE: 19910703
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Feder, Scott B
; REFERENCE/DOCKET NUMBER: 51529
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-372-7842
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS

PCT-US91-04744-3
; Sequence 3, Application PC/TUS9104744
; GENERAL INFORMATION:
; APPLICANT: Vedvick, Thomas S
; APPLICANT: Engel, Michael E
; APPLICANT: Urcan, Mary S
; APPLICANT: Buckholz, Richard G
; APPLICANT: Kinney, Jennifer A
; TITLE OF INVENTION: Production of Aprotinin in
; TITLE OF INVENTION: Methylotrophic Yeast Cells
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 S. LaSalle St.
; CITY: Chicago
; STATE: Ill.
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/04744
; FILING DATE: 19910703
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Feder, Scott B
; REFERENCE/DOCKET NUMBER: 51529
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-372-7842
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; PCT-US91-04744-3

Query Match      8.1%; Score 42.6; DB 6; Length 186;
Best Local Similarity 56.9%; Pred. No. 0.0031;
Matches 78; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 162 tggagctgctatgaagttcacttttagatattctacacagaacctccaaagatgtga 221
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 TGGACCATGCAAGGCCCGTATTATTAGTACTTTTACAACGCTAAGCGCGGACTGTGTCA 89

QY 222 aacttttctctccggctgtaatggaaccttaaacacttcaagcttaaaatagaacg 281
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 88 AACCTTCCTTTACGGTGGATGTAGAGCTAAGAGAAACAACCTCAAGTCTGCTGAGGACTG 29
Qy 282 tgaagtagcctgtgttg 298
Db 28 TATGAGAACCTGTGGTG 12

RESULT 4

5258302-1
; Patent No. 5258302
; APPLICANT: VEDVICK, THOMAS S.; ENGEL, MICHAEL E.; URCAN,
; MARY S.; BUCKHOLZ, RICHARD G.; KINNEY, JENNIFER A.
; TITLE OF INVENTION: DNA FOR EXPRESSION OF APROTININ IN
; METHYLOTROPIC YEAST CELLS
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/560,618
; FILING DATE: 30-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 547,985
; FILING DATE: 03-JUL-1990
; SEQ ID NO: 1:
; LENGTH: 186
5258302-1

Query Match 8.18; Score 42.6; DB 7; Length 186;
Best Local Similarity 56.98; Pred. No. 0.0031;
Matches 78; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
Qy 162 tggagcgtctatgaagttcaacttagatattctacaacagaacctccaaagatgtga 221
Db 39 tggaccatgcagccgcgtattattagatactttacaacgctgaagccgactgtgtca 98
Qy 222 aactttgtcttctccgctgtgaatggcaaccttaacaacttcaagcttaaaatagaacg 281
Db 99 aactttgtttacgctgtgtagtagtagtagtagtagtagtagtagtagtagtagtagtag 158
Qy 282 tgaagtagcctgtgttg 298
Db 159 tatgagaacctgtgtg 175

RESULT 5

US-08-334-773A-5
; Sequence 5, Application US/08334773A
; Patent No. 5629176
; GENERAL INFORMATION:
; APPLICANT: No. 5629176ris, Fanny
; APPLICANT: No. 5629176ris, Kjel
; APPLICANT: Bjorn, Soren Erik
; APPLICANT: Petersen, Lars Christian
; APPLICANT: Olsen, Ole Hvilsted
; TITLE OF INVENTION: A Human Kunitz-Type Protease Inhibitor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56291760 No. 5629176disk of No. 5629176th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/334,773A
; FILING DATE: 04-NOV-1994
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: Agris, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3695.210-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 878 9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: synthetic
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 77..409
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 77..235
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 236..409
; US-08-334-773A-5

Query Match 8.08; Score 42.2; DB 1; Length 418;
Best Local Similarity 53.38; Pred. No. 0.0051;
Matches 89; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
Qy 133 aaagatcctcgaattggacatgaattttggaagctgctatgaagttcaactttatgat 192
Db 239 ACAGATATATCAAGTTTCGCCGAAAGACGAGGAACTTCAGGGATTTCATATTAATGG 298
Qy 193 tctacaacagaaacctccaaagatgaaacttttctctcgcgctgtaagtgaac 252
Db 299 TACTATGATCCAAACACCAAGCTGCAAGATTCTGGTATGGAGGTTGTGGTGGAAAC 358
Qy 253 cttaacaacttcaagcttaaaatagaacgtgaagtagcctgtgttc 299
Db 359 GAAAACAAATTTGGATCACAGAAAGATGTGAAAGGTTTGGCTCC 405

RESULT 6

US-08-084-718-70
; Sequence 70, Application US/08084718
; Patent No. 5591603
; GENERAL INFORMATION:
; APPLICANT: Bjorn, Soren E.
; APPLICANT: No. 5591603ris, Kjel
; APPLICANT: Diness, Viggo
; APPLICANT: No. 5591603rskov-Lauritsen, Lelf
; APPLICANT: Christensen, Niels D.
; APPLICANT: Bregngaard, Claus
; APPLICANT: No. 5591603ris, Fanny
; APPLICANT: Petersen, Lars C.
; TITLE OF INVENTION: Process for the Preparation of Aprotinin
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5591603o No. 5591603disk of No. 5591603th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25


```
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 77..403
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 77..235
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 236..403
;
US-08-084-718-74

Query Match          7.8%; Score 40.8; DB 1; Length 412;
Best Local Similarity 59.5%; Pred. No. 0.012;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 183 ctttagatatttcacacagacgaacctccaaaagatgtgaaacttttctctccggctg 242
Db 283 CATCAGATACTTCTACACGCGCAGGCTGGTTGTGTCAAACTTCGTTTACGGTGGCTG 342

Qy 243 taatggcaaccttaacaaactcaagcttaaaatagaaactggaagtgcctgtgttg 298
Db 343 CAGAGGTAACGGCAACAACCTTCAAGTCTGCTGAAGACTGCATGGAACCTTGTGGTG 398

RESULT 8
US-08-443-976-70
; Sequence 70, Application US/08443976
; Patent No. 5618915
; GENERAL INFORMATION:
; APPLICANT: Bjorn, Soeren E.
; APPLICANT: No. 5618915ris, Kjeld
; APPLICANT: Diness, Viggo
; APPLICANT: No. 5618915rskov-Lauritsen, Lelf
; APPLICANT: Christensen, Niels D.
; APPLICANT: Bregengaard, Claus
; APPLICANT: No. 5618915ris, Fanny
; APPLICANT: Petersen, Lars C.
; TITLE OF INVENTION: Process for the Preparation of Aprotinin
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5618915o No. 5618915disk of No. 5618915th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,976
; FILING DATE: 18-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/084,718
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: DK 4501/87
; FILING DATE: 28-AUG-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2254/88
```

```
; FILING DATE: 26-APR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK88/00138
; FILING DATE: 28-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK89/00096
; FILING DATE: 25-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/466,408
; FILING DATE: 21-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2361/90
; FILING DATE: 01-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/598,337
; FILING DATE: 19-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 1118/91
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00299
; FILING DATE: 01-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,687
; FILING DATE: 29-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/024,925
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3143.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 77..403
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 77..235
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 236..403
;
US-08-443-976-70

Query Match          7.8%; Score 40.8; DB 1; Length 412;
Best Local Similarity 59.5%; Pred. No. 0.012;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 183 ctttagatatttcacacagacgaacctccaaaagatgtgaaacttttctctccggctg 242
Db 283 CATCAGATACTTCTACGAGCGCCACTGCTGGTTGTGTGAAACTTCGTTTACGGTGGCTG 342

Qy 243 taatggcaaccttaacaaactcaagcttaaaatagaaactggaagtgcctgtgttg 298
Db 343 CAGAGGTAACGAGAAACAACCTTCAAGTCTGCTGAAGACTGCATGGAACCTTGTGGTG 398

RESULT 9
US-08-443-976-74
; Sequence 74, Application US/08443976
; Patent No. 5618915
; GENERAL INFORMATION:
; APPLICANT: Bjorn, Soeren E.
```

APPLICANT: No. 5618915ris, Kjeld
APPLICANT: Diness, Viggo
APPLICANT: No. 5618915rskov-Lauritsen, Leif
APPLICANT: Christensen, Niels D.
APPLICANT: Bregengaard, Claus
APPLICANT: No. 5618915ris, Fanny
APPLICANT: Petersen, Lars C.
TITLE OF INVENTION: Process for the Preparation of Aprotinin
TITLE OF INVENTION: and Aprotinin Analogs
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5618915o No. 5618915disk of No. 5618915th America, Inc.
STREET: 405 Lexington Avenue, 62nd Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,976
FILING DATE: 18-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/084,718
FILING DATE: 23-JUN-1993
APPLICATION NUMBER: DK 4501/87
FILING DATE: 28-AUG-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 2254/88
FILING DATE: 26-APR-1988
PRIOR APPLICATION DATA: PCT/DK88/00138
FILING DATE: 28-AUG-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK89/00096
FILING DATE: 25-APR-1989
PRIOR APPLICATION DATA: US 07/466,408
FILING DATE: 21-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 2361/90
FILING DATE: 01-OCT-1990
PRIOR APPLICATION DATA: US 07/598,337
FILING DATE: 19-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 1118/91
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA: PCT/DK91/00299
FILING DATE: 01-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,687
FILING DATE: 29-JAN-1992
PRIOR APPLICATION DATA: US 08/024,925
FILING DATE: 26-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Agriis, Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 3143.224-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 77..403
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 77..235
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 236..403
US-08-443-976-74
Query Match 7.8%; Score 40.8; DB 1; Length 412;
Best Local Similarity 59.5%; Pred. No. 0.012;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 183 ctttagatatttcaacagacaccccaaaagatgtgaaacttttctctcccgctg 242
Db 283 CATCAGATACTTCTACACGCCCAAGGCTGTTGTGTCAAACTTTGCTTTACGGTGGCTG 342
QY 243 taatgcaaccttaacacttcaagcttaaaatagaacgtgaagtagcctgtgttg 298
Db 343 CAGAGGTAACGGCAACACTTCAAGTCTGCTGAAGACTGCATGGAACACTTGTGGTG 398
RESULT 10
US-08-443-977-70
; Sequence 70, Application US/08443977
; Patent No. 5621074
; GENERAL INFORMATION:
; APPLICANT: Bjorn, Soeren E.
; APPLICANT: No. 5621074ris, Kjeld
; APPLICANT: Diness, Viggo
; APPLICANT: No. 5621074rskov-Lauritsen, Leif
; APPLICANT: Christensen, Niels D.
; APPLICANT: Bregengaard, Claus
; APPLICANT: No. 5621074ris, Fanny
; APPLICANT: Petersen, Lars C.
; TITLE OF INVENTION: Process for the Preparation of Aprotinin
; TITLE OF INVENTION: and Aprotinin Analogs
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5621074o No. 5621074disk of No. 5621074th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,977
; FILING DATE: 18-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/084,718
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: DK 4501/87
; FILING DATE: 28-AUG-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2254/88
; FILING DATE: 26-APR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK88/00138
; FILING DATE: 28-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK89/00096
; FILING DATE: 25-APR-1989


```

; APPLICATION NUMBER: US 07/827,687
; FILING DATE: 29-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/024,925
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3143.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 77..409
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 77..235
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 236..409
; US-08-443-977-58

Query Match 7.8%; Score 40.8; DB 1; Length 418;
Best Local Similarity 59.5%; Pred. No. 0.012;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 183 ctttagatattctacaacagaccctccaaaagatgtgaaactttgtctctcccgctg 242
Db 289 CATCAGATACTTCTACGACGCCCACTGCTGGTTGTGTGAAACTTTCGTTTACGGTGGCTG 348

QY 243 taatggcaaccttaacaaacttcaagcttaaaatagaacgtgaagtagcctgtgttg 298
Db 349 CAGAGCTACAGAAACAACTTCAAGTCTGCTGAAGACTGTCATGGAACCTTGTGGTG 404

RESULT 15
US-08-084-718-72
; Sequence 72, Application US/08084718
; Patent No. 5591603
; GENERAL INFORMATION:
; APPLICANT: Bjorn, Soeren E.
; APPLICANT: No. 5591603ris, Kjeld
; APPLICANT: Diness, Viggo
; APPLICANT: No. 5591603rskov-Lauritsen, Leif
; APPLICANT: Christensen, Niels D.
; APPLICANT: Bregengaard, Claus
; APPLICANT: No. 5591603ris, Fanny
; APPLICANT: Petersen, Lars C.
; TITLE OF INVENTION: Process for the Preparation of Aprotinin
; TITLE OF INVENTION: and Aprotinin Analogs
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5591603o No. 5591603disk of No. 5591603th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/084,718
; FILING DATE: 19930623
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 4501/87
; FILING DATE: 28-AUG-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2254/88
; FILING DATE: 26-APR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK88/00138
; FILING DATE: 28-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK89/00096
; FILING DATE: 25-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/466,408
; FILING DATE: 21-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2361/90
; FILING DATE: 01-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/598,337
; FILING DATE: 19-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 1118/91
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00299
; FILING DATE: 01-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,687
; FILING DATE: 29-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/024,925
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3143.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 77..499
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 77..331
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 332..499
; US-08-084-718-72

Query Match 7.5%; Score 39.2; DB 1; Length 508;
Best Local Similarity 58.6%; Pred. No. 0.036;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 183 ctttagatattctacaacagaccctccaaaagatgtgaaactttgtctctcccgctg 242
Db 379 CATCAGATACTTCTACGACGCCCACTGCTGGTTGTGTGAAACTTTCGTTTACGGTGGCTG 438

QY 243 taatggcaaccttaacaaacttcaagcttaaaatagaacgtgaagtagcctgtgttg 298
Db 439 CAGAGCTAAGAGAAACAACTTCAAGTCTGCTGAAGACTGTCATGGAACCTTGTGGTG 494

```

Search completed: May 23, 2000, 23:54:29
Job time: 10381 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:01:00 ; Search time 3550.87 seconds
(without alignments)
-617.577 Million cell updates/sec

Title: US-09-215-435-76
Perfect score: 526
Sequence: 1 ctctgcgtgctgtgtcac.....ctttctgtcaaaaaaaaaa 526

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues
Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA.*
1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq.*

45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US095_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq.*
86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq.*
87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq.*
88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq.*
90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq.*
91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq.*
92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq.*
93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq.*
94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq.*
95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq.*
96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq.*
97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq.*
98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	526	100.0	526	43	US-09-215-435-76 Sequence 76, Appl
2	504.4	95.9	539	43	US-09-215-435-303 Sequence 303, App
3	504.4	95.9	539	70	US-60-069-957-136 Sequence 136, App

```
4 256.4 48.7 271 26 US-08-905-279-148
5 90.2 17.1 208 61 US-09-131-380-189
6 90.2 17.1 208 49 US-60-055-705-189
7 49 9.3 599 48 US-09-308-609-4838
8 49 9.3 599 57 US-09-522-251-4838
9 48.8 9.3 691 49 US-09-306-609-4196
10 48.8 9.3 691 57 US-09-522-251-4196
11 48.2 9.2 437 52 US-09-362-510-19223
12 48.2 9.2 490 46 US-09-289-768-38952
13 48.2 9.2 512 46 US-09-289-768-37096
14 48.2 9.2 512 49 US-09-332-782-30403
15 48.2 9.2 1196 43 US-09-205-070-21132
16 48.2 9.2 1196 50 US-09-340-623-21132
17 48.2 9.2 1412 90 US-60-164-285-4102
18 48.2 9.2 1444 1 PCR-US00-05882-333
19 48.2 9.2 1464 56 US-09-471-275-106
20 48.2 9.2 1464 56 US-09-471-275-692
21 48.2 9.2 1464 56 US-09-491-404-3126
22 48.2 9.2 1464 56 US-09-491-404-3794
23 48.2 9.2 1475 92 US-60-173-373-6986
24 47.8 9.1 336 49 US-09-321-214-20927
25 47.8 9.1 336 57 US-09-516-335-20927
26 47.8 9.1 392 56 US-09-489-036-17177
27 47.8 9.1 396 52 US-09-362-510-17550
28 47.4 9.0 449 48 US-09-306-609-2839
29 47.4 9.0 449 57 US-09-522-251-2839
30 47.2 9.0 534 89 US-60-160-203-3184
31 47.2 9.0 535 89 US-60-160-190-812
32 47.2 9.0 535 89 US-60-160-203-2095
33 46.2 8.8 186 39 US-09-086-253-6
34 46.2 8.8 186 49 US-09-320-095-6
35 46 8.7 153 56 US-09-469-572-3
36 46 8.7 2409 49 US-09-320-095-9
37 45.2 8.6 1313 92 US-60-172-373-18849
38 45.2 8.6 1383 39 US-09-086-253-9
39 45 8.6 471 52 US-09-362-510-18262
40 44.6 8.5 186 39 US-09-086-253-1
41 44.6 8.5 186 49 US-09-320-095-1
42 44.2 8.4 153 56 US-09-469-572-1
43 44.2 8.4 2012 1 PCR-US00-05881-103
44 44 8.4 186 39 US-09-086-253-7
45 44 8.4 186 49 US-09-320-095-7
```

ALIGNMENTS

```
RESULT 1
US-09-215-435-76
; Sequence 76, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 76
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: 22..318
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 22..93
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.6
; OTHER INFORMATION: seq PFIFCSLNTLLG/GV
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 497..502
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 516..526
; US-09-215-435-76

Query Match 100.0%; Score 526; DB 43; Length 526;
Best Local Similarity 100.0%; Pred. No. 1.8e-125;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcctgctgcttgcaccatgaagtctgcacacgctggaattcttcttaagattcttc 60
Db 1 ctgcctgctgcttgcaccatgaagtctgcacacgctggaattcttcttaagattcttc 60
QY 61 atcttctgctcattgaataccctgttattgggtgggtgttaataaaaaattcggagaagata 120
Db 61 atcttctgctcattgaataccctgttattgggtgggtgttaataaaaaattcggagaagata 120
QY 121 tctggagacctcaaaagatccctgcacaaattggacatgaatttgggaagctgctatgaagtt 180
Db 121 tctggagacctcaaaagatccctgcacaaattggacatgaatttgggaagctgctatgaagtt 180
QY 181 cactttagatattctcacacagaaacctcaaaagatgtgaaacttttctctccggc 240
Db 181 cactttagatattctcacacagaaacctcaaaagatgtgaaacttttctctccggc 240
QY 241 tgtaatggcaaccttaacaaacttcaagcttaaaatagaacgtaagctgctgttgca 300
Db 241 tgtaatggcaaccttaacaaacttcaagcttaaaatagaacgtaagctgctgttgca 300
QY 301 aaatacaaacaccgagggtgagaggatgtgaactatgaagttgtctgtgacccatccg 360
Db 301 aaatacaaacaccgagggtgagaggatgtgaactatgaagttgtctgtgacccatccg 360
QY 361 aaataaagacacagaataatcagactgatttgaacttcttgaataatttccataatgc 420
Db 361 aaataaagacacagaataatcagactgatttgaacttcttgaataatttccataatgc 420
QY 421 tttaagcttccatgatttggctattttctcctgaccctagtttcttcttcttggaattaa 480
Db 421 tttaagcttccatgatttggctattttctcctgaccctagtttcttcttcttggaattaa 480
QY 481 ctgtatgatacttagaataaagagctcttctctcaaaaaa 526
Db 481 ctgtatgatacttagaataaagagctcttctctcaaaaaa 526
```

RESULT 2

```
US-09-215-435-303
; Sequence 303, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
```

; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 303
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 32..328
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 32..103
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.59999990463257
; OTHER INFORMATION: seq FFIFCSLNTLLG/GV
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 508..513
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 528..539
US-09-215-435-303

Query Match 95.9%; Score 504.4; DB 43; Length 539;
Best Local Similarity 96.8%; Pred. No. 6.7e-120;
Matches 511; Conservative 12; Mismatches 3; Indels 2; Gaps 1;

```
Qy . 1 ctgctgctgctgctgacacatgaagctgccaagctggagattcttctaagattcttc 60
Db 11 ctgctgctgctgctgacacatgaagctgccaagctggagattcttctaagattcttc 70

Qy 61 atcttctgctcattgaatacctgttattgggtgtgttaataaaattgcggaagata 120
Db 71 atcttctgctcattgaatacctgttattgggtgtgttaataaaattgcggaagata 130

Qy 121 tgtggagacctcaagaatccctgcaaatgtggacatgaatttggaaagctgctatgaagt 180
Db 131 tgtggagacctcaagaatccctgcaaatgtggacatgaatttggaaagctgctatgaagt 190

Qy 181 cactttagatatctcaacagacaccccaagaatgtgaaacctttgtcttctccggc 240
Db 191 cactttagatatctcaacagacaccccaagaatgtgaaacctttgtcttctccaggc 250

Qy 241 tgaatggcaaccttaacacacctcaagcttaaaatagaacgtgaagctgtgttgca 300
Db 251 tgaatggcaaccttaacacacctcaagcttaaaatagaacgtgaagctgtgttgca 310

Qy 301 aaatacaaacaccaggtgagaggatgtgaactcatgaagttgtctgctgcaccatccg 360
Db 311 aaatacaaacaccaggtgagaggatgtgaactcatgaagttgtctgctgcaccatccg 370

Qy 361 aaatacaaacaccaggtgagaggatgtgaactcatgaagttgtctgctgcaccatccg 420
Db 371 aaatacaaacaccaggtgagaggatgtgaactcatgaagttgtctgctgcaccatccg 430

Qy 420 -cttaagcttcccatgtgtgctatttctgacctagtttggcttcttctcctggaatt 478
Db 431 cgwtaagcttcccatgtgtgctatttctgacctagtttggcttcttctcctggaatt 490

Qy 479 aactgtatgatcatagataaagaagcttcttctgcaaaaaaa 526
Db 491 aactgtatgatcatagataaagaagcttcttctgcaaaaaaa 538
```

RESULT 3
US-60-069-957-136
; Sequence 136, Application US/60069957
; GENERAL INFORMATION:

; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 381
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/069,957
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.019PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 136:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 539 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Testis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 32..328
; NAME/KEY: sig_peptide
; LOCATION: 32..103
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 4.6
; OTHER INFORMATION: seq FFIFCSLNTLLG/GV
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 508..513
; FEATURE:
; NAME/KEY: polyA
; LOCATION: 528..539
; IDENTIFICATION METHOD: blastn
; US-60-069-957-136

Query Match 95.9%; Score 504.4; DB 70; Length 539;
Best Local Similarity 96.8%; Pred. No. 6.7e-120;
Matches 511; Conservative 12; Mismatches 3; Indels 2; Gaps 1;

```
Qy 1 ctgctgctgctgctgacacatgaagctgccaagctggagattcttctaagattcttc 60
Db 11 ctgctgctgctgctgacacatgaagctgccaagctggagattcttctaagattcttc 70

Qy 61 atcttctgctcattgaatacctgttattgggtgtgttaataaaattgcggaagata 120
Db 71 ATCTCTCTCTCATTAATACCTCTGTTATGGGTGGTGTATATAAAATTCGCGAAGATA 130

Qy 121 tgtggagacctcaagaatccctgcaaatgtggacatgaatttggaaagctgctatgaagt 180
Db 131 TGTGGAGACCTCAAAGATCCCTGCAAAATTTGGACATGAATTTTGGAAAGTGTATGAAGTT 190

Qy 181 cactttagatatctcaacagacaccccaagaatgtgaaacctttgtcttctccggc 240
```

Db 191 CACTTAGATATTTACAAACAGACCTCAAAAGATGTGAACATTTTGTCTCTCCAGC 250
QY 241 tgaatggcaacttaacaactcaagcttaaaataagaacgtgaagtgtgtgca 300
Db 251 TGAATGGCAACTTAAACATTTCAAGCTTAAATAAGACGTGAAGTAKCCTGTGTGCA 310
QY 301 aaatacaaaacccaggggtgagaggatggaactcatgaagtgtgtctgtgcaccatcgg 360
Db 311 AAATACAAACCCGAGGTGAGAGATGTGAACCTATGAGTTGTCTGTGCACCATCCG 370
QY 361 aaataagacacacagaataatcgaactgattttgaaatcttttgttaatttccataatg- 420
Db 371 AAATAAGACACACAGAAATTCARACTGATTTWGAATCTTTTATTTCCMYNAKGG 430
QY 420 -cttaagcttcacatgtttgtctatttctcgaacctgattttgtcttctctggaaatt 478
Db 431 CGWRTAAGCTTCCATATGTTTGTATTTTCTGACCTAGTTTGTCTTCTCGGAAATT 490
QY 479 aactgtatgatcatagaaatgaagagctcttctgtcgaataaaaaaaa 526
Db 491 AACTGTATGAKATTASATGAAGAGTCTTTCTGTCAAAAAAANA 538

RESULT 4

US-08-905-279-148
; Sequence 148, Application US/08905279
; GENERAL INFORMATION:

; APPLICANT: Genset SA

; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS

; TITLE OF INVENTION: EXPRESSED IN TESTIS AND OTHER TISSUES

; NUMBER OF SEQUENCES: 503

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knorbe, Martens, Olson & Bear

; STREET: 501 West Broadway

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92101-3505

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: Win95

; SOFTWARE: Word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/905,279

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Israelsen, Ned A.

; REGISTRATION NUMBER: 29,655

; REFERENCE/DOCKET NUMBER: GENSET.011A

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 148:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 271 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: DOUBLE

; TOPOLOGY: LINEAR

; MOLECULE TYPE: CDNA

; ORIGINAL SOURCE:

; ORGANISM: Homo Sapiens

; TISSUE TYPE: Testis

; FEATURE:

; NAME/KEY: sig_peptide

; LOCATION: 32..103

; IDENTIFICATION METHOD: Von Heijne matrix

; OTHER INFORMATION: score 4.6

; OTHER INFORMATION: seq FFIFCSLNTLLG/GV

US-08-905-279-148

Query Match 48.7%; Score 256.4; DB 26; Length 271;
Best Local Similarity 99.6%; Pred. No. 3.7e-56;
Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ctgctgtctgtctgtgcaccatgaagtctgcagctgggattttctctasgattcttc 60
Db 11 CTGCTGTCTGTCTGTGCACCATGAAGTCTGCAAGCTGGGATTTCTTCAAGATTCTTC 70
QY 61 atctctgtcattgaataaccctgtatttggtgtgttaataaaattcgcgagaagata 120
Db 71 ACTTCTGTCTCAATTGAATACCTCTGTTATTGGGTGGTGTAAATAAATTCGGGAGAAGATA 130
QY 121 tftggagacctcaaaagatccctgcgaattggacatgaattttggagctgctatgaagtt 180
Db 131 TGTGGAGACCTCAAGATCCCTGCAATTTGGACATGAATTTTGGAGAGCTGCTATGAAGTT 190
QY 181 cactttagattttctacaacagaaacctcaaaagatggaaccttttctctctccggc 240
Db 191 CACTTTAGATATTCTTACAAACAGAACCTCCAAAGATGTGAACCTTTTGTCTCTCCAGC 250
QY 241 tgaatggcaaccttaac 258
Db 251 TGTATGGCAACCTTAAC 268

RESULT 5

US-09-131-380-189

; Sequence 189, Application US/09131380

; GENERAL INFORMATION:

; APPLICANT: Gooding, Douglas H.

; APPLICANT: Stuve, Laura L.

; APPLICANT: Stuart, Susan G.

; APPLICANT: Naughton, Rebecca E.

; APPLICANT: Garrow, Bonnie L.

; APPLICANT: Klemm, Juli D.

; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM RAT FEMUR

; FILE REFERENCE: PZ-0013 US

; CURRENT APPLICATION NUMBER: US/09/131,380

; CURRENT FILING DATE: 1998-08-07

; EARLIER APPLICATION NUMBER: 60/055,705

; EARLIER FILING DATE: 1997-08-08

; EARLIER APPLICATION NUMBER: 60/074,278

; EARLIER FILING DATE: 1998-02-10

; NUMBER OF SEQ ID NOS: 2199

; SOFTWARE: PERL Program

; SEQ ID NO 189

; LENGTH: 208

; TYPE: DNA

; ORGANISM: Rattus norvegicus

; FEATURE:

; NAME/KEY: unsure

; LOCATION: 109, 117, 156, 167

; OTHER INFORMATION: a or g or c or t, unknown, or other

; FEATURE: -

; OTHER INFORMATION: 700292496H1

US-09-131-380-189

Query Match 17.1%; Score 90.2; DB 41; Length 208;
Best Local Similarity 67.1%; Pred. No. 2e-13;
Matches 139; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

QY 97 gtaataaaaattcggagagatattggagacctcaagatccctgcgaattgggacatg 156
Db 2 gtgagagaaggttgctgcgaaggctctgcaagaattcaacgacctctgtgttggaag 61
QY 157 aatttggagctgctatgaagtccacttttagattttctacaacagaaacctccaa-aa 215
Db 62 gatcctggcagctgctatgaagccattacagattttctacaacacacctccanccaa 121
QY 216 atgtgaaacttttctctccggtgtaatggcaaccttaacacctcaagcttaaaat 275


```
; NAME/KEY: misc_feature
; LOCATION: (1)...(599)
; OTHER INFORMATION: n = A,T,C or G
US-09-522-251-4838

Query Match          9.3%; Score 49; DB 57; Length 599;
Best Local Similarity 57.0%; Pred. No. 0.012;
Matches 85; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggacatgaatttggagcgtgctatgaagttcacttttagatatt 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 aagattcctgcagctgggtactcgcgcgtccctgcgtggaatgaccagcaggtatt 269
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 194 tctacaacagaaacctccaaaagatgtgaaacttttcttctcgcgctgtaattggcaacc 253
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 tctataatggtacatccatgacctgtgagactttccagtcnccgctgcatgggnaacg 329
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 254 ttaacaacttcaagcttaaaatagaacgt 282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 gtaacaacttcntcacagaaaaggantgt 358
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-09-306-609-4196
; Sequence 4196, Application US/09306609
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-718CON2
; CURRENT APPLICATION NUMBER: US/09/306,609
; CURRENT FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 09/168,296
; EARLIER FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: US 08/949,553
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 13025
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4196
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(691)
; OTHER INFORMATION: n = A,T,C or G
US-09-306-609-4196

Query Match          9.3%; Score 48.8; DB 48; Length 691;
Best Local Similarity 57.7%; Pred. No. 0.015;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggacatgaatttggagcgtgctatgaagttcacttttagatatt 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 295 aagattcctgcagctgggtactcgcgcgtccctgcgtggaatgaccancaggtatt 354
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 194 tctacaacagaaacctccaaaagatgtgaaacttttcttctcgcgctgtaattggcaacc 253
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 355 tctataatggtacatccatgacctgtgagactttccagtcacgctgctgctgggcaacg 414
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 254 ttaacaacttcaagcttaaaatagaacgt 282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 gtaacaacttcgtcacagaaaaggantgt 443
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-09-362-510-19223
; Sequence 19223, Application US/09362510
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-759CON1
; CURRENT APPLICATION NUMBER: US/09/362,510
; CURRENT FILING DATE: 1999-07-27
; EARLIER APPLICATION NUMBER: US 09/221,820
; EARLIER FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 62165
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19223
; LENGTH: 437
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-362-510-19223

Query Match          9.2%; Score 48.2; DB 52; Length 437;
Best Local Similarity 57.7%; Pred. No. 0.016;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggacatgaatttggagcgtgctatgaagttcacttttagatatt 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7 aagattcctgccagctgggtactcgcgcgtccctgcgtggaatgaccagcaggtatt 66
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 194 tctacaacagaaacctccaaagaatgtgaaacttttcttctccggctgtaatggcaacc 253
      ||||| || || ||||| ||||| ||||| || ||||| || ||||| || |||||
Db 67 tctataatgttacatccatcgctgtgagactttccagtaacggcgctcatgggcaacg 126

QY 254 ttaacaacttcaagcttaaaatagaacgt 282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 gtaacaacttcgtcacagaaaaggagtgt 155

RESULT 12
US-09-289-768-38952
; Sequence 38952, Application US/09289768
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-765
; CURRENT APPLICATION NUMBER: US/09/289,768
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 39996
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 38952
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(490)
; OTHER INFORMATION: n = A,T,C or G
US-09-289-768-38952

Query Match          9.2%; Score 48.2; DB 46; Length 490;
Best Local Similarity 57.7%; Pred. No. 0.018;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggacatgaattttggaagctgtatgaagttcactttagatatt 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82 aagattcttcagctgggtactcggcggtccctgcatacgggaatgaccagcaggattt 141

QY 194 tctacaacagaaacctccaaagaatgtgaaacttttcttctccggctgtaatggcaacc 253
      ||||| || || ||||| ||||| ||||| || ||||| || |||||
Db 142 tctataatgttacatccatcgctgtgagactttccagtaacggcgctcatgggcaacg 201

QY 254 ttaacaacttcaagcttaaaatagaacgt 282
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 gtaacaacttcgtcacagaaaaggagtgt 230

RESULT 13
US-09-289-768-37096
; Sequence 37096, Application US/09289768
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-765
; CURRENT APPLICATION NUMBER: US/09/289,768
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 39996
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37096
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(512)
; OTHER INFORMATION: n = A,T,C or G
US-09-289-768-37096
```

```
Query Match          9.2%; Score 48.2; DB 46; Length 512;
Best Local Similarity 57.7%; Pred. No. 0.019;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggacatgaattttggaagctgtatgaagttcactttagatatt 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 aagattcttcagctgggtactcggcggtccctgcatacgggaatgaccagcaggattt 215

QY 194 tctacaacagaaacctccaaagaatgtgaaacttttcttctccggctgtaatggcaacc 253
      ||||| || || ||||| ||||| ||||| || ||||| || |||||
Db 216 tctataatgttacatccatcgctgtgagactttccagtaacggcgctcatgggcaacg 275

QY 254 ttaacaacttcaagcttaaaatagaacgt 282
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 gtaacaacttcgtcacagaaaaggagtgt 304

RESULT 14
US-09-332-782-30403
; Sequence 30403, Application US/09332782A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/332,782A
; CURRENT FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30403
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(512)
; OTHER INFORMATION: n = A,T,C or G
US-09-332-782-30403

Query Match          9.2%; Score 48.2; DB 49; Length 512;
Best Local Similarity 57.7%; Pred. No. 0.019;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 134 aagatccctgcaaatggacatgaattttggaagctgtatgaagttcactttagatatt 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 aagattcttcagctgggtactcggcggtccctgcatacgggaatgaccagcaggattt 215

QY 194 tctacaacagaaacctccaaagaatgtgaaacttttcttctccggctgtaatggcaacc 253
      ||||| || || ||||| ||||| ||||| || ||||| || |||||
Db 216 tctataatgttacatccatcgctgtgagactttccagtaacggcgctcatgggcaacg 275

QY 254 ttaacaacttcaagcttaaaatagaacgt 282
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 gtaacaacttcgtcacagaaaaggagtgt 304

RESULT 15
US-09-205-070-21132
; Sequence 21132, Application US/09205070
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-748
; CURRENT APPLICATION NUMBER: US/09/205,070
; CURRENT FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21132
; LENGTH: 1196
; TYPE: DNA
; ORGANISM: Homo sapiens
```

US-09-205-070-21132

Query Match 9.2% Score 48.2; DB 43; Length 1196;
Best Local Similarity 57.7%; Pred. No. 0.025;
Matches 86; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
Qy 134 aagatccctgcaaatggacatgaatttggaaagctgctatgaagttcacttttagatatt 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 829 aagattcctgcagctgggctactcgcgcggtccctgcagtggaatgaccagcaggtatt 888
Qy 194 tctacaacagaacctccaaaagatgtgaaacttttgccttcctccggctgtaatggcaacc 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 889 tctataatggtacatccatgacctgtgagactttccagtagcggcggtgcagggcaacg 948
Qy 254 ttaacaacttcaagcttaaaatagaacgt 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 949 gtaacaacttcgtcacagaaaaggagtgt 977

Search completed: May 24, 2000, 01:01:04
Job time: 14236 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:10:18 ; Search time 6115.43 Seconds
(without alignments)
-86.217 Million cell updates/sec

Title: US-09-215-435-78

Perfect score: 542

Sequence: 1 cagcactgtggccatgat.....aaaaaaaaaaaaaaaaaaaaa 542

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_bal.*
2: gb_ba2.*
3: gb_ba2.*
4: gb_ov.*
5: gb_ov.*
6: gb_ph.*
7: gb_ph.*
8: gb_pl1.*
9: gb_pl1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_ro.*
14: gb_sy.*
15: gb_sy.*
16: gb_v1.*
17: gb_v1.*
18: em_hum1.*
19: em_hum2.*
20: em_hum2.*
21: em_hum2.*
22: em_or.*
23: em_or.*
24: em_pat.*
25: em_pat.*
26: em_pl.*
27: em_ro.*
28: em_ro.*
29: em_sy.*
30: em_sy.*
31: em_v1.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba1.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*
51: gb_pr5.*
52: gb_htg8.*
53: gb_htg9.*
54: gb_htg10.*
55: gb_htg11.*
56: gb_htg12.*
57: gb_htg13.*
58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	48.4	8.9	452	3	CCOLIP
2	46.6	8.6	167439	44	AC020737
3	44.8	8.3	490	12	MCCOLIPAS
4	42.6	7.9	459	3	RABCOLIPAS
5	41.4	7.6	118150	11	HSJ800C24
6	41.2	7.6	163720	33	CNS01DXB
7	41.2	7.6	171918	32	CNS01DUN
8	40.6	7.5	497	3	ECOLIPB
9	40.4	7.5	446	12	RATCOLIP
10	40.4	7.5	492	12	RATCOLQ
11	40.4	7.5	1195	12	RATNCAMC
12	40.4	7.5	227070	56	AC023888
13	40	7.4	206784	10	HS93L7
14	39.4	7.3	183341	43	AC011974
15	39.4	7.3	188651	42	AC010092
16	39.2	7.2	187750	41	AC009455
17	39	7.2	748	11	HSM801177
18	39	7.2	145295	55	AC021778
19	38.8	7.2	42301	9	AP000528
20	38.8	7.2	149939	32	CNS01DVK
21	38.8	7.2	153449	33	HSJ730HL6
22	38.8	7.2	172798	11	CNS01DVF
23	38.8	7.2	204338	32	AL136532
24	38.8	7.2	231559	42	AC010102
25	38.6	7.1	55699	44	AC021157
26	38.6	7.1	68777	45	AC021620
27	38.4	7.1	166782	56	AC008812
28	38.2	7.0	152408	10	HS191J18
29	38	7.0	7218	5	I66494
30	38	7.0	102456	9	AP000083
31	38	7.0	121127	45	AC013795
32	38	7.0	194347	33	AL137069
33	37.8	7.0	111008	42	AC010061
34	37.8	7.0	135686	10	HS473J6
35	37.8	7.0	158251	44	AC018783
36	37.8	7.0	165533	33	AL139819
37	37.8	7.0	173324	44	AC009131
38	37.8	7.0	176248	33	AL133255
39	37.8	7.0	207264	33	AL138921
40	37.8	7.0	232528	33	HS28124
41	37.8	7.0	232528	33	HS28124
42	37.8	7.0	311844	33	AL137015
43	37.6	6.9	62493	10	HS193G15
44	37.6	6.9	62493	10	HSJ193G15
45	37.6	6.9	81506	56	AC023809

ALIGNMENTS

```
RESULT 1
CCOLIP          452 bp  mRNA          MAM          31-MAR-1995
LOCUS          Canine mRNA for colipase.
DEFINITION
ACCESSION      X53564
VERSION        X53564.1  GI:841
KEYWORDS       colipase.
SOURCE         dog.
ORGANISM       Canis familiaris
REFERENCE      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS        Carnivora; Fissipedia; Canidae; Canis.
TITLE          1 (bases 1 to 452)
JOURNAL        Fukuoka S.I. and Scheele, G.A.
                Direct Submission
                Submitted (19-JUN-1990) Fukuoka S.-I., Scheele G.A., The Charles A.
                Dana Research Institute, Harvard Medical School, Cell and Molecular
                Biology, Dana 533, Beth Israel Hospital, 330 Brookline Ave.,
                Boston, MA 02215, USA
REFERENCE      2 (bases 1 to 452)
AUTHORS        Fukuoka S., Taniguchi, Y., Kitagawa, Y. and Scheele, G.
TITLE          Full length cDNA sequence encoding canine pancreatic colipase
JOURNAL        Nucleic Acids Res. 18 (18), 5549 (1990)
MEDLINE        91016846
FEATURES       Location/Qualifiers
source         1..452
               /organism="Canis familiaris"
               /db_xref="taxon:9615"
               /dev_stage="adult"
               /tissue_type="pancreas"
               /clone_lib="pUC-9"
               /clone="pCL1"
sig_peptide    1..51
               /product="signal peptide"
CDS            1..339
               /codon_start=1
               /product="precursor peptide"
               /protein_id="CAA37636.1"
               /db_xref="GI:842"
               /translation="WEKILVLLVALVYVDPRIIHLDELCLNSVOCKSK
               CHRVGLSLARCAPRASENSESAKTLGVYKPCERGLTCGDKSIVGSIITNFG
               VCHDAGRSKK"
mat_peptide    52..336
misc_feature    429..435
               /product="colipase"
polya_site     452
               /note="polyA signal"
               /note="polyA site"
BASE COUNT     90 a 146 c 117 g 99 t
ORIGIN
Query Match      8.9%; Score 48.4; DB 3; Length 452;
Best Local Similarity 49.6%; Pred. No. 0.0017;
Matches 124; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 39 gctgctgttcctctctctctctctctctctctctctctccaccaggggctcactttctccaac 98
DB 9 GATCCCTCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 68
QY 99 aaaaataaaccttttggagctcaaggagtgcttgcattcccggaaccagagctcgagactgg 158
DB 69 GATCATATATCCACTGGAAGAGCGGCGAGCTCTGTCTGAACAGCGTCCAGTGAAGAGCAA 128
QY 159 ctgctgccaactgtccagacaattgcgagtcgactgcgggagagaaggggtccgaggg 218
DB 129 GTGCTGCCACCGGGGCCACCGGGGTGAGCTGGCCCGCTGGCCACCAAGCCAGCCAGAGAA 188
QY 219 cagctgtgtcacaacgaggtgtctctttggccaataagacgctgccccctgctgcggaa 278
DB 189 CAGCGAGTGTCTGCAACACGCTCTATGGGGTTTACTACAGTGTCTCTGTGAGCGGGG 248
```

```
QY 279 cctgacttgt 288
DB 249 CCTGACCTGT 258

RESULT 2
AC020737
LOCUS          Homo sapiens clone RP11-75009, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
ACCESSION      AC020737
VERSION        AC020737.1  GI:6682696
KEYWORDS       HTG; HTGS_PHASE0.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          1 (bases 1 to 167439)
JOURNAL        Waterston, R.H.
                The sequence of Homo sapiens clone
                Unpublished
REFERENCE      2 (bases 1 to 167439)
AUTHORS        Waterston, R.H.
TITLE          Direct Submission
JOURNAL        Submitted (08-JAN-2000) Genome Sequencing Center, Washington
                University School of Medicine, 444 Forest Park Parkway, St. Louis,
                MO 63108, USA
COMMENT        Center project name: H.NH075009.
                * NOTE: This record contains 235 individual
                * sequencing reads that have not been assembled into
                * contigs. Runs of N are used to separate the reads
                * and the order in which they appear is completely
                * arbitrary. Low-pass sequence sampling is useful for
                * identifying clones that may be gene-rich and allows
                * overlap relationships among clones to be deduced.
                * However, it should not be assumed that this clone
                * will be sequenced to completion. In the event that
                * the record is updated, the accession number will
                * be preserved.
                * 1 668: contig of 668 bp in length
                * 669 678: gap of unknown length
                * 679 1184: contig of 506 bp in length
                * 1185 1508: contig of 314 bp in length
                * 1509 1518: gap of unknown length
                * 1519 1792: contig of 274 bp in length
                * 1793 1802: gap of unknown length
                * 1803 2476: contig of 674 bp in length
                * 2477 2486: gap of unknown length
                * 2487 3155: contig of 669 bp in length
                * 3156 3166: gap of unknown length
                * 3811 3820: gap of unknown length
                * 3821 4484: contig of 664 bp in length
                * 4485 4494: gap of unknown length
                * 4495 5130: contig of 636 bp in length
                * 5131 5140: gap of unknown length
                * 5141 6200: contig of 1060 bp in length
                * 6201 6210: gap of unknown length
                * 6211 6878: contig of 668 bp in length
                * 6879 6888: gap of unknown length
                * 6889 7727: contig of 839 bp in length
                * 7728 7737: gap of unknown length
                * 7738 7838: contig of 661 bp in length
                * 8408 8409: gap of unknown length
                * 8409 8850: contig of 442 bp in length
                * 8851 8860: gap of unknown length
                * 8861 9627: contig of 767 bp in length
                * 9628 9637: gap of unknown length
                * 9638 10282: contig of 645 bp in length
                * 10283 10292: gap of unknown length
                * 10293 10957: contig of 665 bp in length
                * 10958 10967: gap of unknown length
                * 10968 11635: contig of 668 bp in length
```

```
* 11636: gap of unknown length
* 11645: contig of 924 bp in length
* 12569: gap of unknown length
* 12579: contig of 1218 bp in length
* 13797: gap of unknown length
* 13807: contig of 664 bp in length
* 13808: gap of unknown length
* 14472: contig of 679 bp in length
* 14482: gap of unknown length
* 15161: contig of 233 bp in length
* 15171: gap of unknown length
* 15404: contig of 719 bp in length
* 15414: gap of unknown length
* 16132: contig of 659 bp in length
* 16143: gap of unknown length
* 16801: contig of 298 bp in length
* 16812: gap of unknown length
* 17119: contig of 669 bp in length
* 17120: gap of unknown length
* 17788: contig of 670 bp in length
* 17798: gap of unknown length
* 18468: contig of 668 bp in length
* 18478: gap of unknown length
* 18479: contig of 669 bp in length
* 19147: gap of unknown length
* 19156: contig of 669 bp in length
* 19825: gap of unknown length
* 19835: contig of 679 bp in length
* 19836: gap of unknown length
* 20514: contig of 664 bp in length
* 20524: gap of unknown length
* 21188: contig of 652 bp in length
* 21189: gap of unknown length
* 21850: contig of 642 bp in length
* 21860: gap of unknown length
* 21861: contig of 1035 bp in length
* 22709: gap of unknown length
* 22718: contig of 673 bp in length
* 23483: gap of unknown length
* 23493: contig of 682 bp in length
* 23494: gap of unknown length
* 24176: contig of 759 bp in length
* 24177: gap of unknown length
* 24819: contig of 662 bp in length
* 24828: gap of unknown length
* 25863: contig of 800 bp in length
* 25873: gap of unknown length
* 25884: contig of 481 bp in length
* 26569: gap of unknown length
* 26570: contig of 672 bp in length
* 27635: gap of unknown length
* 27644: contig of 665 bp in length
* 27964: gap of unknown length
* 27974: contig of 497 bp in length
* 28656: gap of unknown length
* 28666: contig of 620 bp in length
* 29425: gap of unknown length
* 29435: contig of 662 bp in length
* 29436: gap of unknown length
* 30566: contig of 680 bp in length
* 30576: gap of unknown length
* 31096: contig of 662 bp in length
* 31106: gap of unknown length
* 31168: contig of 662 bp in length
* 31778: gap of unknown length
* 31789: contig of 680 bp in length
* 32458: gap of unknown length
* 32468: contig of 481 bp in length
* 32949: gap of unknown length
* 32959: contig of 717 bp in length
* 33676: gap of unknown length
* 33686: contig of 665 bp in length
* 34351: gap of unknown length
* 34361: contig of 497 bp in length
* 34362: gap of unknown length
* 34858: gap of unknown length
* 34869: contig of 672 bp in length
* 35541: gap of unknown length
* 35551: contig of 313 bp in length
* 35864: gap of unknown length
* 35873: contig of 465 bp in length
* 36338: gap of unknown length
* 36348: contig of 465 bp in length
```

```
* 36349: contig of 655 bp in length
* 37004: gap of unknown length
* 37014: contig of 671 bp in length
* 37685: gap of unknown length
* 37694: contig of 690 bp in length
* 38384: gap of unknown length
* 38395: contig of 406 bp in length
* 38800: gap of unknown length
* 38810: contig of 679 bp in length
* 38811: gap of unknown length
* 39490: contig of 294 bp in length
* 39499: gap of unknown length
* 39793: contig of 294 bp in length
* 39794: gap of unknown length
* 39804: contig of 294 bp in length
* 40097: gap of unknown length
* 40107: contig of 653 bp in length
* 40108: gap of unknown length
* 40760: contig of 666 bp in length
* 40771: gap of unknown length
* 41436: contig of 669 bp in length
* 41446: gap of unknown length
* 41447: contig of 669 bp in length
* 42115: gap of unknown length
* 42125: contig of 256 bp in length
* 42381: gap of unknown length
* 42391: contig of 677 bp in length
* 43068: gap of unknown length
* 43078: contig of 661 bp in length
* 43079: gap of unknown length
* 43740: contig of 654 bp in length
* 43750: gap of unknown length
* 44404: contig of 477 bp in length
* 44414: gap of unknown length
* 44891: contig of 693 bp in length
* 44891: gap of unknown length
* 45593: contig of 694 bp in length
* 45603: gap of unknown length
* 45604: contig of 643 bp in length
* 46297: gap of unknown length
* 46298: contig of 680 bp in length
* 46308: gap of unknown length
* 46951: contig of 640 bp in length
* 46961: gap of unknown length
* 47641: contig of 233 bp in length
* 47651: gap of unknown length
* 48290: contig of 670 bp in length
* 48291: gap of unknown length
* 48301: contig of 670 bp in length
* 48533: gap of unknown length
* 48534: contig of 472 bp in length
* 48544: gap of unknown length
* 49213: contig of 1125 bp in length
* 49224: gap of unknown length
* 49224: contig of 520 bp in length
* 49695: gap of unknown length
* 49696: contig of 953 bp in length
* 49706: gap of unknown length
* 50831: contig of 664 bp in length
* 50841: gap of unknown length
* 51360: contig of 664 bp in length
* 51371: gap of unknown length
* 51371: contig of 668 bp in length
* 52266: gap of unknown length
* 52277: contig of 671 bp in length
* 53229: gap of unknown length
* 53230: contig of 1034 bp in length
* 53240: gap of unknown length
* 53903: contig of 668 bp in length
* 53904: gap of unknown length
* 53914: contig of 668 bp in length
* 54582: gap of unknown length
* 54592: contig of 671 bp in length
* 55263: gap of unknown length
* 55273: contig of 1034 bp in length
```

Query Match 8.6%; Score 46.6; DB 44; Length 167439;
Best Local Similarity 59.4%; Pred. No. 0.0084;
Matches 79; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy 410 cctccacctgtctctccctaccagagctctgttccacctgttccctccagagctcc 469

Db 149824 CCC 149883

Oy 470 accatgagtggagggaagtggggagtgattgaaataaagagcttttttcaatgaaaaaa 529

Db 149884 CCCCCCGCGGG 149943

```
QY 530 aaaaaaaaaa 542
|||||
Db 149944 AAAAAAAAAA 149956

RESULT 3
MCCOLIPAS
LOCUS MCCOLIPAS 490 bp mRNA 21-NOV-1996
DEFINITION M.coypus mRNA for colipase.
ACCESSION X82998
VERSION X82998.1 GI:599866
KEYWORDS colipase.
SOURCE nutria.
ORGANISM Myocastor coypus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Hystricognathi; Myocastoridae; Myocastor.
1 (bases 1 to 490)
Thirstrup,K., Carriere,F., Hjorth,S.A., Rasmussen,P.B.,
Nielsen,P.F., Ladefoged,C., Thim,L. and Boel,E.
Cloning and expression in insect cells of two pancreatic lipases
and a procolipase from Myocastor coypus
Eur. J. Biochem. 227 (1-2), 186-193 (1995)
JOURNAL 95154288
MEDLINE 2 (bases 1 to 490)
REFERENCE Thirstrup,K.
AUTHORS Direct Submission
TITLE Submitted (28-NOV-1994) K. Thirstrup, Univ. Hospital, Lab. for
JOURNAL Molecular Endocrinology, Rigshospitalet 6321, Blegdamsvej 9,
DK-2100 Copenhagen, DENMARK
FEATURES
source
1..490
Location/Qualifiers
/organism="Myocastor coypus"
/db_xref="taxon:10157"
/tissue_type="pancreas"
46..381
/codon_start=1
/product="colipase"
/protein_id="CAA58119.1"
/db_xref="GI:599867"
/db_xref="SWISS-PROT:P42889"
/translation="MEKVLVLLFLVAVAYADPRGLIINLDNGELCLNSAQCKSQ
COHDSPLGLARCAADKARENSGCPQTIYIYLCPCERGLTCDGDKSIIGAITNTNYG
ICQDPQSKK"
46..96
sig_peptide
mat_peptide 97..378
/note="putative propeptide: 97..111"
/misc_feature
97..111
/note="putative propeptide"
/product="colipase"
BASE COUNT 96 a 181 c 120 g 93 t
ORIGIN

Query Match 8.3%; Score 44.8; DB 12; Length 490;
Best Local Similarity 53.4%; Pred. No. 0.019;
Matches 94; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 113 tggagctcaaggagttgttcacccgaacaggactgcgagactggctgtccaaactgtg 172
|||||
Db 128 TGGACAACGGCGAGCTCTCCTGAACAGCGCCCGAGTGCAAGAGTCAAGTCTGCCAGCAG 187
|||||
QY 173 ctccagacaattgcgagctgcgactgcgagagaagggtcccgaggcagctgtgtcaca 232
|||||
Db 188 ACAGCCCCCTGGGCGCTGGCCGCTGGCCGACACAGCCACAGACAGCGGCTGTCTCC 247
|||||
QY 233 cgcaggtgtttctttggcccaatagacgtgtccctgcctgcggaaacctgtgt 288
|||||
Db 248 CACAGACGATCTATGGAATCTACTACTTGTGTCCTGTCCCTGCGAGCGGCGCTGTG 303
|||||

RESULT 4
RABCOLIPAS
LOCUS RABCOLIPAS 459 bp mRNA MAM 05-OCT-1993
DEFINITION Oryctolagus cuniculus pancreatic colipase mRNA, complete cds.
ACCESSION L06329
VERSION L06329.1 GI:164893
KEYWORDS colipase.
SOURCE Oryctolagus cuniculus
ORGANISM Oryctolagus cuniculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Lagomorpha; Leporidae; Oryctolagus.
1 (bases 1 to 459)
Colwell,N.S., Aleman-Gomez,J.A., Sasser,T. and Kumar,V.B.
Cloning and characterization of rabbit pancreatic colipase
Int. J. Biochem. 25 (6), 885-890 (1993)
JOURNAL 93345715
MEDLINE
FEATURES
source
1..459
Location/Qualifiers
/organism="Oryctolagus cuniculus"
/db_xref="taxon:9986"
39..89
sig_peptide
CDS 39..362
/codon_start=1
/product="colipase"
/protein_id="AAA02911.1"
/db_xref="GI:164894"
/translation="MEKVLVLLVALSVAYAAPGRGIVINLEGEGLCLNSAQCKSGC
CHHSALSARCAPKASENSECSPTIYGVYKPCERGLTCEGDKSIVGSIYVNTNFG
VCLDV"
90..359
mat_peptide
BASE COUNT 84 a 170 c 122 g 83 t
ORIGIN

Query Match 7.9%; Score 42.6; DB 3; Length 459;
Best Local Similarity 48.2%; Pred. No. 0.083;
Matches 120; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 39 gctgtcttcctctctctctctctctctctctctctctcaccagggtctcactttctccaac 98
|||||
Db 47 GGTCTTGTCTGCTGCTGTGTGGCCCTCTCTGGTGCTTACGAGCTCTCTGGCCCCGGGG 106
|||||
QY 99 aaaaataaccttttggagctcaaggagtcttgcattcccgaaaccaggactgcgagactgg 158
|||||
Db 107 AATCGTTATCAACTGGAGGAGGGGAGCTCTGCTTGAACAGCGCCAGTGAAGAGCGG 166
|||||
QY 159 ctgctgccaactgtctccagacaattgcgagctgcgactgcgagagaagggtccagagg 218
|||||
Db 167 CTGCTGCCACCACTCGAGCGCTCTGAGTCTGGCCGCTGCCACCCAGCCAGCGAGAA 226
|||||
QY 219 cagtctgttcaaacgcaggtgttctttggccaatatagagcgtgtccctgcctgcggaa 278
|||||
Db 227 CAGCGAGTGTTCGCCCCAGACCATCTACGCGCTCTACTACAAGTGTCCCTGTGAGCGCG 286
|||||
QY 279 cctgacttg 287
|||||
Db 287 CCTGACCTG 295

RESULT 5
HSJ800C24
LOCUS HSJ800C24 118150 bp DNA PRI 15-DEC-1999
DEFINITION Human DNA sequence from clone RP4-800C24 on chromosome 20, complete
sequence.
ACCESSION AL121593
VERSION AL121593.10 GI:6523758
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 118150)
REFERENCE Blakey,S.
AUTHORS
```


ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
TITLE	Eutheria; Primates; Catarrhini; Homidae; Homo.
JOURNAL	1 (bases 1 to 227070)
REFERENCE	Waterston, R.H.
TITLE	The sequence of Homo sapiens clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 227070)
TITLE	Waterston, R.H.
JOURNAL	Submitted (18-FEB-2000) Genome Sequencing Center, Washington
COMMENT	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	* NOTE: This record contains 369 individual
COMMENT	* sequencing reads that have not been assembled into
COMMENT	* contigs. Runs of N are used to separate the reads
COMMENT	* and the order in which they appear is completely
COMMENT	* arbitrary. Low-pass sequence sampling is useful for
COMMENT	* identifying clones that may be gene-rich and allows
COMMENT	* overlap relationships among clones to be deduced.
COMMENT	* However, it should not be assumed that this clone
COMMENT	* will be sequenced to completion. In the event that
COMMENT	* the record is updated, the accession number will
COMMENT	* be preserved.
COMMENT	* 1
COMMENT	* 504: contig of 504 bp in length
COMMENT	* 505
COMMENT	* 1026: contig of 512 bp in length
COMMENT	* 1037
COMMENT	* 1036: gap of unknown length
COMMENT	* 1298: contig of 262 bp in length
COMMENT	* 1308: gap of unknown length
COMMENT	* 1309
COMMENT	* 2144: contig of 836 bp in length
COMMENT	* 2154: gap of unknown length
COMMENT	* 2155
COMMENT	* 2627: contig of 473 bp in length
COMMENT	* 2628
COMMENT	* 3093: gap of unknown length
COMMENT	* 3094
COMMENT	* 3103: gap of unknown length
COMMENT	* 3613: contig of 510 bp in length
COMMENT	* 3614
COMMENT	* 3623: gap of unknown length
COMMENT	* 3624
COMMENT	* 4071: contig of 448 bp in length
COMMENT	* 4072
COMMENT	* 4081: gap of unknown length
COMMENT	* 4082
COMMENT	* 4418: gap of unknown length
COMMENT	* 4409
COMMENT	* 4419
COMMENT	* 4930: contig of 512 bp in length
COMMENT	* 4931
COMMENT	* 4940: gap of unknown length
COMMENT	* 4941
COMMENT	* 5388: contig of 448 bp in length
COMMENT	* 5389
COMMENT	* 5398: gap of unknown length
COMMENT	* 5797
COMMENT	* 5806: gap of unknown length
COMMENT	* 6322: contig of 516 bp in length
COMMENT	* 6323
COMMENT	* 6332: gap of unknown length
COMMENT	* 6333
COMMENT	* 6848: contig of 516 bp in length
COMMENT	* 6849
COMMENT	* 7372: contig of 514 bp in length
COMMENT	* 6859
COMMENT	* 7373
COMMENT	* 7382: gap of unknown length
COMMENT	* 7383
COMMENT	* 7896: contig of 514 bp in length
COMMENT	* 7897
COMMENT	* 7906: gap of unknown length
COMMENT	* 7907
COMMENT	* 8420: contig of 514 bp in length
COMMENT	* 8430
COMMENT	* 8431: contig of 401 bp in length
COMMENT	* 8832
COMMENT	* 8841: gap of unknown length
COMMENT	* 9352: contig of 511 bp in length
COMMENT	* 9353
COMMENT	* 9362: gap of unknown length
COMMENT	* 9858: contig of 496 bp in length
COMMENT	* 9363
COMMENT	* 9868: gap of unknown length
COMMENT	* 9859
COMMENT	* 10380: contig of 512 bp in length
COMMENT	* 9869
COMMENT	* 10381
COMMENT	* 10390: gap of unknown length
COMMENT	* 10391
COMMENT	* 10785: contig of 395 bp in length
COMMENT	* 10786
COMMENT	* 10795: gap of unknown length
COMMENT	* 10796
COMMENT	* 10897: contig of 102 bp in length
COMMENT	* 10898
COMMENT	* 10907: gap of unknown length
COMMENT	* 10908
COMMENT	* 11422: contig of 515 bp in length
COMMENT	* 11423
COMMENT	* 11432: gap of unknown length
COMMENT	* 11433
COMMENT	* 12251: contig of 819 bp in length
COMMENT	* 12252
COMMENT	* 12261: gap of unknown length
COMMENT	* 12772
COMMENT	* 12772: contig of 511 bp in length
COMMENT	* 12782: gap of unknown length
COMMENT	* 12783
COMMENT	* 13262: contig of 480 bp in length
COMMENT	* 13263
COMMENT	* 13272: gap of unknown length
COMMENT	* 13602: contig of 330 bp in length
COMMENT	* 13603
COMMENT	* 13612: gap of unknown length
COMMENT	* 13613
COMMENT	* 14125: contig of 513 bp in length
COMMENT	* 14126
COMMENT	* 14135: gap of unknown length
COMMENT	* 14136
COMMENT	* 14643: contig of 508 bp in length
COMMENT	* 14644
COMMENT	* 14653: gap of unknown length
COMMENT	* 15125: contig of 472 bp in length
COMMENT	* 15126
COMMENT	* 15135: gap of unknown length
COMMENT	* 15136
COMMENT	* 15646: contig of 511 bp in length
COMMENT	* 15647
COMMENT	* 15656: gap of unknown length
COMMENT	* 16169: contig of 513 bp in length
COMMENT	* 16170
COMMENT	* 16179: gap of unknown length
COMMENT	* 16180
COMMENT	* 16691: contig of 512 bp in length
COMMENT	* 16692
COMMENT	* 16701: gap of unknown length
COMMENT	* 16702
COMMENT	* 17213: contig of 512 bp in length
COMMENT	* 17214
COMMENT	* 17223: gap of unknown length
COMMENT	* 17224
COMMENT	* 17734: contig of 511 bp in length
COMMENT	* 17735
COMMENT	* 17744: gap of unknown length
COMMENT	* 18571: contig of 827 bp in length
COMMENT	* 18572
COMMENT	* 18581: gap of unknown length
COMMENT	* 18582
COMMENT	* 19414: contig of 833 bp in length
COMMENT	* 19415
COMMENT	* 19424: gap of unknown length
COMMENT	* 19937: contig of 513 bp in length
COMMENT	* 19938
COMMENT	* 19947: gap of unknown length
COMMENT	* 19948
COMMENT	* 20457: contig of 510 bp in length
COMMENT	* 20458
COMMENT	* 20467: gap of unknown length
COMMENT	* 20468
COMMENT	* 20978: contig of 511 bp in length
COMMENT	* 20979
COMMENT	* 20988: gap of unknown length
COMMENT	* 21498: contig of 510 bp in length
COMMENT	* 21499
COMMENT	* 21508: gap of unknown length
COMMENT	* 21509
COMMENT	* 22021: contig of 513 bp in length
COMMENT	* 22022
COMMENT	* 22031: gap of unknown length
COMMENT	* 22032
COMMENT	* 22540: contig of 509 bp in length
COMMENT	* 22541
COMMENT	* 22550: gap of unknown length
COMMENT	* 22551
COMMENT	* 23394: contig of 844 bp in length
COMMENT	* 23395
COMMENT	* 23404: gap of unknown length
COMMENT	* 23405
COMMENT	* 24229: contig of 825 bp in length
COMMENT	* 24230
COMMENT	* 24239: gap of unknown length
COMMENT	* 24240
COMMENT	* 24752: contig of 513 bp in length
COMMENT	* 24753
COMMENT	* 24762: gap of unknown length
COMMENT	* 24839: contig of 77 bp in length
COMMENT	* 24840
COMMENT	* 24849: gap of unknown length
COMMENT	* 24850
COMMENT	* 25362: contig of 513 bp in length
COMMENT	* 25363
COMMENT	* 25372: gap of unknown length
COMMENT	* 25373
COMMENT	* 25885: contig of 513 bp in length
COMMENT	* 25886
COMMENT	* 25895: gap of unknown length
COMMENT	* 25896
COMMENT	* 26407: contig of 512 bp in length
COMMENT	* 26408
COMMENT	* 26417: gap of unknown length
COMMENT	* 26418
COMMENT	* 26931: contig of 514 bp in length
COMMENT	* 26932
COMMENT	* 26941: gap of unknown length
COMMENT	* 26942
COMMENT	* 27136: contig of 195 bp in length
COMMENT	* 27137
COMMENT	* 27146: gap of unknown length
COMMENT	* 27338: contig of 192 bp in length
COMMENT	* 27339
COMMENT	* 27348: gap of unknown length
COMMENT	* 27349
COMMENT	* 27564: contig of 216 bp in length
COMMENT	* 27565
COMMENT	* 27574: gap of unknown length
COMMENT	* 27833: contig of 259 bp in length
COMMENT	* 27834
COMMENT	* 27843: gap of unknown length
COMMENT	* 28355: contig of 512 bp in length
COMMENT	* 28356
COMMENT	* 28365: gap of unknown length
COMMENT	* 28366
COMMENT	* 28876: contig of 511 bp in length
COMMENT	* 28877
COMMENT	* 28886: gap of unknown length
COMMENT	* 29400: contig of 514 bp in length
COMMENT	* 29401
COMMENT	* 29410: gap of unknown length
COMMENT	* 29924: contig of 514 bp in length
COMMENT	* 29925
COMMENT	* 29934: gap of unknown length
COMMENT	* 30744: contig of 810 bp in length

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:59:41 ; Search time 282.38 Seconds
(without alignments)
480.219 Million cell updates/sec

Title: US-09-215-435-78

Perfect score: 542

Sequence: 1 cagcacctgtggccatgat.....aaaaaaaaaaaaaaaaaaaaa 542

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	361.8	66.8	421	X51465	Human secreted pro
2	36.4	6.7	275	Q50196	RGAL/3'-end. Rham
3	36.4	6.7	1326	X00633	Human secreted pro
4	36	6.6	704	V35167	Thuga plicata diri
5	35.2	6.5	972	X04341	Human secreted pro
6	35.2	6.5	1142	X27314	Human secreted pro
7	35	6.5	1308	X22114	Human secreted pro
8	35	6.5	2835	T72666	Cell cycle checkpo
9	35	6.5	8239	T61981	Human ataxia and r
10	34.8	6.4	700	V34301	Human secreted pro
11	34.8	6.4	3641	V03351	Homo sapiens inter
12	34.6	6.4	1605	V59687	Human secreted pro
13	34.4	6.3	1003	X08984	Sequence encoding
14	34.4	6.3	3623	X82821	Rabbit osteoclast-
15	34	6.3	604	X37373	Human secreted pro
16	34	6.3	3201	T66232	Rat huntingtin ass
17	34	6.3	3708	T66231	Rat huntingtin ass
18	34	6.3	29392	V15422	Mouse poly Ig rece
19	33.8	6.2	1041	V08823	Gene No. 13 encodi
20	33.8	6.2	2496	V63189	CDNA from clone cr
21	33.6	6.2	867	T02135	Human secreted pro
22	33.6	6.2	867	T88057	Partial cDNA clone
23	33.6	6.2	2734	Q39221	Encodes human epid
24	33.6	6.2	2734	Q47376	HETG mRNA. Diagnos
25	33.4	6.2	635	T075445	Murine glycosylati
26	33.4	6.2	635	T38390	Murine glycosylati
27	33.4	6.2	1147	V69620	Human secreted pro
28	33.4	6.2	2184	Q06301	Sequence encoding
29	33.2	6.1	245	V30925	Human secreted pro
30	33.2	6.1	1817	X16674	Human secreted pro
31	33.2	6.1	1041	Q54651	T84.12 light chain
32	33.2	6.1	1041	Q54653	T84.12 L4-12-1 lig
33	33.2	6.1	2911	V84579	Human secreted pro
34	33	6.1	698	X15448	CDNA encoding a mu

RESULT 1

X51465
ID X51465 standard; cDNA; 421 BP.
AC X51465;
DT 21-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:44.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906549-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1231.
PR 01-AUG-1997; US-905279.
PA (GENSET) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153779/13.
DR P-PSDB; Y12687.
PT New nucleic acids encoding human secreted proteins - obtained from
cDNA libraries derived from testis, ovary, uterus and spleen tissue
PS Claim 1; Page 177; 522pp; English.
CC X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human
secreted proteins, and encode the proteins given in Y12681 to Y12913,
respectively. The proteins given represent the signal peptide and an
N-terminal fragment of a secreted protein. The nucleic acid sequences
can be used for producing secreted human gene products. They can also
be used to develop products for diagnosis and therapy. The proteins
obtained may have cytokine activity, cell proliferation/differentiation
activity, haematopoiesis regulating activity, tissue growth regulating
activity, reproductive hormone regulating activity, chemotactic/
chemokinetic activity, haemostatic and thrombolytic activity, receptor/
ligand activity, anti-inflammatory activity, tumour inhibition activity
or other activities. The products can be used in forensic, gene therapy
and chromosome mapping procedures. The sequences can also be used for
obtaining corresponding promoter sequences. The nucleic acids encoding
the signal peptide can be used for directing extracellular secretion of
a polypeptide or the insertion of a polypeptide into a membrane, or
importing a polypeptide into a cell.
SQ Sequence 421 BP; 96 A; 106 C; 110 G; 102 T;

Query Match 66.8%; Score 361.8; DB 1; Length 421;

Best Local Similarity 97.6%; Pred. No. 2.7e-89;

Matches 360; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cagcacctgtggccatgatctaccaccaatggctgctgctgctgtctctctctcttc 60

Db 47 CACGACCTGTGGCCATGCTACCCCAATGGCTGCTGCTGCTCTCTCTCTTC 106

Qy 61 ttctctctctctctacacagggggtcactttctcaacaaataaaccttttgagctc 120

Db 107 TTTCTCTCTCTCTCTCACACAGGGGCTCACTTTCTCAACAAATACACGCTTTTGAGCTC 166

Qy 121 aaggagctgttcacccggaacacgagactgcgagagctgctgctgcacacgctgcacac 180

```
Db 167 AAGAGSKRKGATSGGGAACACAGGACTCGGAGACTGGGTGCTGCCAGCTGCTCCAGAC 226
QY 181 aatgagagtcgcaactgcgcgagaaaggggtccgagggcagtcgtgtgtaaacgcaggtg 240
Db 227 AATTGCGAGTCGCACTGCGGAGAGAGGGTCCGAGGGCAGTCTGTGTCAACGCGAGGTG 286
QY 241 ttcttggccaatagagcgtgtccctgcctgcgcggaacctgacttgatatattcaaaag 300
Db 287 TTCTTTGGCAATATAGAGCGTGTCCCTGCTCGGAACCTGACTTGTATATATCAAG 346
QY 301 aatgagaaatgcttagcatcgctcctgtgcccgtgtgcagaaatggaagcagaggtg 360
Db 347 AATGAGAAATGGCTTAGCATCGCTATGCGCCGTGTGCAGAAATTTGGAAGGCGAGAGTTG 406
QY 361 gctaagaaa 369
Db 407 GCTAGRAAA 415

RESULT 2
ID Q50196
AC Q50196 standard; DNA; 275 BP.
DT 04-MAY-1994 (first entry)
DE RGA1/3'-end.
KW Rhamnogalacturonan acetyl esterase; RGAE; enzyme; Aspergillus aculeatus;
KW modified hairy region; MHR; probe; pYES 2.0; amplification; primer;
KW polymerase chain reaction; PCR; ss.
OS Aspergillus aculeatus RE4.
FH Key Location/Qualifiers
FT 1..162
FT /*tag= a
FT /note= "C-terminal"
FT
FT WO9320190-A.
PN 14-OCT-1993.
PD 29-MAR-1993. DK0109.
PF 27-MAR-1992. DK-000420.
PR (NOVO) NOVO-NORDISK AS.
PI Christensen FM, Dalboge H, Dorreich K, Heldt-Hansen HP,
PI Mischler M, Schnell Y;
PI WPI; 93-336898/42.
DR P-PSDB; R43265.
DT Rhamno-galacturonan acetyl esterase enzyme obtd. from Aspergillus
PT aculeatus - used to degrade acetylated modified hairy region of
PT vegetable material
PS Disclosure; Fig 9; 58pp; English.
CC The probe and primers (Q50192-94) were used in the amplification
CC of RGAE DNA from Aspergillus aculeatus RE4. The 5' and 3'-end
CC sequences are given in Q50195-96.
SQ Sequence 275 BP; 90 A; 51 C; 81 G; 53 T;

Query Match 6.7%; Score 36.4; DB 1; Length 275;
Best Local Similarity 70.0%; Pred. No. 0.46;
Matches 49; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 473 atgagtgaggagaaagtgaggagtgatgaataaagagcgtttttcaatgaaaaa 532
Db 199 AGGAGTGGAGGGGTAAGGACAGGAGGTGTTCTATGTATAGTATCGACAAAAA 258
QY 533 aaaaaaaa 542
Db 259 AAAAAAAA 268

RESULT 3
ID X00633
AC X00633 standard; DNA; 1326 BP.
DT 25-MAR-1999 (first entry)
DE Human secreted protein gene 23 clone HEKFL13.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
```

diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; developmental abnormality; foetal deficiency; blood; allergy; renal; ds; immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS; cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; osteoporosis; arthritis; testis; lung; thyroditis; thyroid; digestion; endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.
PN WO9842738-A1.
PD 01-OCT-1998. U05311.
PF 19-MAR-1998; US-050937.
PR 30-MAY-1997; US-041276.
PR 21-MAR-1997; US-041277.
PR 21-MAR-1997; US-041281.
PR 21-MAR-1997; US-042344.
PR 30-MAY-1997; US-048069.
PR 30-MAY-1997; US-048094.
PR 30-MAY-1997; US-048095.
PR 30-MAY-1997; US-048096.
PR 30-MAY-1997; US-048099.
PR 30-MAY-1997; US-048131.
PR 30-MAY-1997; US-048135.
PR 30-MAY-1997; US-048154.
PR 30-MAY-1997; US-048160.
PR 30-MAY-1997; US-048186.
PR 30-MAY-1997; US-048187.
PR 30-MAY-1997; US-048188.
PR 30-MAY-1997; US-048350.
PR 30-MAY-1997; US-048351.
PR 30-MAY-1997; US-048352.
PR 30-MAY-1997; US-048355.
PR 05-AUG-1997; US-054804.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA,
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,
PI Rosen CA, Ruben SM, Shi Y, Young P;
DR WPI; 99-070066/06.
DR P-PSDB; W67829.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 189; 385pp; English.
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the Descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X00602) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic acid
CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 87 polynucleotides, based on
CC which tissues they are most highly expressed in (see X00611 for described
CC uses).

QY 439 ctctgttccacctgttccccagagcctccaccatgagtgaggaggaagtggggagtgat 498
Db 1209 CTGTCTTCTCCCGGGTTCACAGCTCCGACCCCTCGCCCATGAAGGAGCTGGCAGGT 1268
QY 499 tgaataaagagctttttcaatgaaaaaataaaaaa 542
Db 1269 GGAATAAACAACAACACTTTATTAAAAA 1312

Query Match 6.7%; Score 36.4; DB 1; Length 1326;
Best Local Similarity 58.7%; Pred. No. 0.83;
Matches 61; Conservative 1; Mismatches 42; Indels 0; Gaps 0;

Db 554 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 591

Search completed: May 23, 2000, 23:59:52
Job time: 10678 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:54:29 ; Search time 203.6 Seconds
(without alignments)
346.031 Million cell updates/sec

Title: US-09-215-435-78
Perfect score: 542
Sequence: 1 cagcactgtggccatgat.....aaaaaaaaaaaaaaaaaaaa 542

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 450926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	7.0	7218	1 US-08-232-463-14	Sequence 14, Appl
2	34.4	6.3	1003	4 US-08-887-997B-1	Sequence 1, Appl
3	33.4	6.2	635	2 US-08-455-633A-35	Sequence 35, Appl
4	33.4	6.2	635	2 US-08-416-336-5	Sequence 5, Appl
5	33.4	6.2	635	4 US-08-456-460C-35	Sequence 35, Appl
6	33.4	6.2	635	6 PCT-US94-05354-35	Sequence 35, Appl
7	33	6.1	2069	1 US-07-921-178A-1	Sequence 1, Appl
8	32.8	6.1	361	3 US-08-465-380-39	Sequence 39, Appl
9	32.8	6.1	361	3 US-08-486-397-39	Sequence 39, Appl
10	32.8	6.1	361	3 US-08-486-399-39	Sequence 39, Appl
11	32.8	6.1	361	3 US-08-461-965-39	Sequence 39, Appl
12	32.8	6.1	361	4 US-08-634-641-39	Sequence 39, Appl
13	32.8	6.1	361	5 US-09-249-471-39	Sequence 39, Appl
14	32.8	6.1	642	5 US-08-946-026-33	Sequence 33, Appl
15	32.8	6.1	3214	1 US-08-484-105-17	Sequence 17, Appl
16	32.8	6.1	3214	1 US-08-484-106-17	Sequence 17, Appl
17	32.6	6.0	1386	4 US-08-897-340-1	Sequence 1, Appl
18	32.4	6.0	547	1 US-08-131-365B-41	Sequence 41, Appl
19	32.4	6.0	547	3 US-08-668-123-41	Sequence 41, Appl
20	32.4	6.0	722	2 US-08-698-551-5	Sequence 5, Appl
21	32.4	6.0	722	3 US-08-602-228-5	Sequence 5, Appl
22	32.4	6.0	722	3 US-08-649-341A-5	Sequence 5, Appl
23	32.4	6.0	722	3 US-08-494-440B-5	Sequence 5, Appl
24	32.4	6.0	722	3 US-08-533-901B-5	Sequence 5, Appl
25	32.4	6.0	722	3 US-08-839-032A-5	Sequence 5, Appl
26	32.4	6.0	722	4 US-08-839-031A-5	Sequence 5, Appl
27	32.4	6.0	722	6 PCT-US95-12724-5	Sequence 5, Appl

28 32.2 5.9 289 1 US-08-341-568-3 Sequence 3, Appl
29 32.2 5.9 289 3 US-08-911-020-3 Sequence 3, Appl
30 32.2 5.9 318 2 US-08-702-344-17 Sequence 17, Appl
31 32.2 5.9 2125 1 US-07-779-890-3 Sequence 3, Appl
32 32.2 5.9 2125 2 US-07-779-890-3 Sequence 3, Appl
33 32.2 5.9 2125 6 PCT-US93-05640-3 Sequence 3, Appl
34 32.2 5.9 2589 6 PCT-US96-12860-1 Sequence 1, Appl
35 32 5.9 247 2 US-08-691-641-4 Sequence 4, Appl
36 32 5.9 372 3 US-08-630-822A-96 Sequence 96, Appl
37 32 5.9 372 4 US-09-005-069-96 Sequence 96, Appl
38 32 5.9 654 6 PCT-US95-06406A-11 Sequence 11, Appl
39 32 5.9 1365 1 US-08-114-072-1 Sequence 1, Appl
40 32 5.9 1365 6 PCT-US94-09361-1 Sequence 1, Appl
41 31.8 5.9 1868 2 US-08-658-883B-1 Sequence 1, Appl
42 31.8 5.9 2823 1 US-08-398-008A-1 Sequence 1, Appl
43 31.8 5.9 2823 4 US-08-893-333-1 Sequence 1, Appl
44 31.6 5.8 827 6 PCT-US95-06406A-20 Sequence 20, Appl
45 31.6 5.8 1293 1 US-08-317-522A-6 Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pt29pt-Fls
US-08-232-463-14

RESULT 3

```

US-08-455-633A-35
; Sequence 35, Application US/08455633A
; Patent No. 5786168
; GENERAL INFORMATION:
; APPLICANT: Ishizaka, Kimishige
; APPLICANT: Liu, Yun-Cai
; APPLICANT: Mikayama, Toshifumi
; TITLE OF INVENTION: METHOD FOR RECOMBINANT PRODUCTION OF
; BIOLOGICALLY ACTIVE POLYPEPTIDES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455.633A
; FILING DATE: 31-May-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07246/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: murine GIF
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 82..426
;
US-08-455-633A-35

Query Match 6.2%; Score 33.4; DB 2; Length 635;
Best Local Similarity 54.5%; Pred. No. 0.72; Mismatches 56; Indels 0; Gaps
Matches 67; Conservative 0;

QY 420 ctctcctccaccagagctctgtttcacccctgttccccagagcgtccaccatgagtg 479
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 CTGCGCTCCACGTAAGTGTCTGTGTTATCCACCGGTAGCGATGCCACCTTCAGCC 527
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 480 gaggggaagtggggagtgattgaataaagagtgcttttccaatgaaaaa 539
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 GGGAGAAATAAATGGTTTATAGAGACCAAAAAAAAAAAAAAAAAAAAAA 587
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 540 aaa 542
|||
Db 588 AAA 590

RESULT 4
US-08-416-336-5
; Sequence 5, Application US/08416336
; Patent No. 5807714
; GENERAL INFORMATION:
; APPLICANT: Ishizaka, Kimishige
; APPLICANT: Ishii, Yasuyuki
; TITLE OF INVENTION: METHOD OF PRODUCTION OF ANTIGEN-SPECIFIC

```

Query Match 6.28; Score 33.4; DB 4; Length 635;

Best Local Similarity 34.5%, Freq. NO: 0.72,
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy	420	ctctctccctaccagagctctgtgttaacctgtgttccccagagctccaacctgagt	479
Db	468	CTGCCTCTCCAGTAGTGTCTGTGTTATTCACCGGTAGCGATGCCACTTCCAGCC	527
Qy	480	gagggagagtggggagtgattgaaataaagagcttttccaatgaaaaaataaaaaaaa	539
Db	528	GGGAGAGAATAAATGGTTTATAGAGACCAAAAAAAAAAAAAAAAAAAAAAAAAA	587
Qy	540	aaa	542
Db	588	AAA	590

RESULT 6

```

PCT-US94-05354-35
; Sequence 35, Application PC/TUS9405354
; GENERAL INFORMATION:
; APPLICANT: La Jolla Institute for
; APPLICANT: Allergy and Immunology
; TITLE OF INVENTION: METHOD FOR RECOMBINANT PRODUCTION OF
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE POLYPEPTIDES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05354
; FILING DATE: 13-MAY-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., Ph.D., John R.
; REGISTRATION NUMBER: 31,678

```

RESULT 5

```

US-08-456-460C-35
; Sequence 35, Application US/08456460C
; Patent No. 5945096
; GENERAL INFORMATION:
; APPLICANT: Ishizaka, Kimishige
; APPLICANT: Liu, Yun-Cai
; APPLICANT: Mikayama, Toshifumi
; TITLE OF INVENTION: METHOD FOR RECOMBINANT PRODUCTION OF
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE POLYPEPTIDES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; REFERENCE/DOCKET NUMBER: FD-2581
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: murine GIF
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 82..426
; PCT-US94-05354-35

Query Match 6.2%; Score 33.4; DB 6; Length 635;
Best Local Similarity 54.5%; Pred. No. 0.72;
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 420 ctctctccctaccagagctctgtttccaccctgtttcccccagagcctccaccatgagt 479
Db 468 CTCGCCCTCCACGTAGTGTCTGTGTTTATCCACCGGTAGCGTCCACCTTCAGCC 527
Qy 480 gaggaagtggggagtgattgaataaagagctttttcaatgaataaataaataa 539
Db 528 GGGAGAAATAATGGTTATTAAGAGACCAAAAAAATAAATAAATAAATAAATAA 587
Qy 540 aaa 542
Db 588 AAA 590

RESULT 7
US-07-921-178A-1
; Sequence 1, Application US/07921178A
; Patent No. 5356775
; GENERAL INFORMATION:
; APPLICANT: Hebert, Steven C
; APPLICANT: Ho, Kevin
; TITLE OF INVENTION: A CDNA SEQUENCE ENCODING A NOVEL FAMILY
; TITLE OF INVENTION: OF ATP SENSITIVE POTASSIUM CHANNEL, AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921,178A
; FILING DATE: 19920729
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FOX, SAM L
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0627.3250000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 466-0800
; TELEFAX: (202) 833-8716
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2069 base pairs
; TYPE: NUCLEIC ACID

; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 150..1322
; US-07-921-178A-1

Query Match 6.1%; Score 33; DB 1; Length 2069;
Best Local Similarity 54.5%; Pred. No. 1.8;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 422 ctctctccctaccagagctctgtttccaccctgtttcccccagagcctccaccatgagt 481
Db 1948 CTGAGCCCATACAGAAATTTGGCCTGAATGATTTCCCTGTGGAGCATTAATGGAGGCCAA 2007
Qy 482 ggggaagtggggagtgattgaataaagagcctttttcaatgaataaataaataa 541
Db 2008 CTCACACTCTTTAGATATTAATGAATATCTTTTGCAGAGAAAAAATAAATAA 2067
Qy 542 a 542
Db 2068 A 2068

RESULT 8
US-08-465-380-39
; Sequence 39, Application US/08465380
; Patent No. 5863894
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Berquim
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESS: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,380
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```


Query Match	6.1%	Score 32.8;	DB 3;	Length 361;
Best Local Similarity	64.5%	Pred. No. 0.79;		
Matches	49: Conservative	0: Mismatches	27: Indels	

```

; ORIGINAL SOURCE:
; ORGANISM: Necator americanus

```

```
;
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16....252
US-08-634-641-39

Query Match          6.1%; Score 32.8; DB 4; Length 361;
Best Local Similarity 64.5%; Pred. No. 0.79;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Oy 457 tccaccatgagtggagggaagtggagggtgattgaaataaagagctttttcaatgaaaaa 526
Db 277 tcttacattagatgggtgagctgattctgtcagataaaactctttcttcttaaaaaaa 336
Oy 527 aaaaaaaaaaaaaa 542
Db 337 aaaaaaaaaaaaaa 352

RESULT 13
US-09-249-471-39
; Sequence 39, Application US/09249471
; Patent No. 6040441
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; FILING DATE:
; PRIOR APPLICATION NUMBER: US/09/249,471
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCI/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
```

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16....252
US-09-249-471-39

Query Match          6.1%; Score 32.8; DB 5; Length 361;
Best Local Similarity 64.5%; Pred. No. 0.79;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Oy 457 tccaccatgagtggagggaagtggagggtgattgaaataaagagctttttcaatgaaaaa 526
Db 277 tcttacattagatgggtgagctgattctgtcagataaaactctttcttcttaaaaaaa 336
Oy 527 aaaaaaaaaaaaaa 542
Db 337 aaaaaaaaaaaaaa 352

RESULT 14
US-08-946-026-33
; Sequence 33, Application US/08946026
; Patent No. 6034218
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Mitcham, Jennifer L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,026
; FILING DATE: 07-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 642 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-946-026-33
```

Search completed: May 23, 2000, 23:54:38
Job time: 10390 sec

Query Match 6.1%; Score 32.8; DB 5; Length 642;
Best Local Similarity 76.9%; Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 491 ggagtgattgaagaagcgttttcaatgaaaaaaaaaaaaaaaaaaaaa 542
||| | ||||| || | ||||| ||||| ||||| ||||| ||||| |||||
Db 586 GGAATAATCCATAATAAGGATATTCATTATAAAAAAAAAAAAAAAAAA 637

RESULT 15
US-08-484-105-17
; Sequence 17, Application US/08484105
; Patent No. 5589341
; GENERAL INFORMATION:
; APPLICANT: STILLMAN, Bruce
; APPLICANT: BELL, Stephen P
; APPLICANT: KOBAYASHI, Ryuji
; APPLICANT: RINE, Jasper
; APPLICANT: FOSS, Margit
; APPLICANT: McNALLY, Francis J
; APPLICANT: LAURENSEN, Patricia
; APPLICANT: HERSKOWITZ, Ira
; APPLICANT: LI, Joachim J
; APPLICANT: GAVIN, Kimberly
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/484,105
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3214 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 220..2802
US-08-484-105-17

Query Match 6.1%; Score 32.8; DB 1; Length 3214;
Best Local Similarity 84.1%; Pred. No. 2.6;
Matches 37; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 499 tgaataaagcgttttcaatgaaaaaaaaaaaaaaaaaaaaa 542
||| ||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db 3135 TTAATAATTGCTTTTAAAAAAAAAAAAAAAAAAAAAAAAA 3178

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:01:04 ; Search time 3550.87 Seconds
(without alignments)
-636.363 Million cell updates/sec

Title: US-09-215-435-78

Perfect score: 542

Sequence: 1 cagcagctgtggccatgat.....aaaaaaaaaaaaaaaaaaaa 542

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues

Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending_Patents_NA:*

- 1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
- 2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
- 3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
- 4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
- 5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq:*
- 6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq:*
- 7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq:*
- 8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq:*
- 9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq:*
- 10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq:*
- 11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq:*
- 12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq:*
- 13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq:*
- 14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq:*
- 15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq:*
- 16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq:*
- 17: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
- 18: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
- 19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq:*
- 20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq:*
- 21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq:*
- 22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq:*
- 23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq:*
- 24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq:*
- 25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq:*
- 26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq:*
- 27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq:*
- 28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq:*
- 29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq:*
- 30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq:*
- 31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq:*
- 32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq:*
- 33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq:*
- 34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq:*
- 35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq:*
- 36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq:*
- 37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq:*
- 38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq:*
- 39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq:*
- 40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq:*
- 41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq:*
- 42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq:*
- 43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq:*
- 44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq:*

- 45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq:*
- 46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq:*
- 47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq:*
- 48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq:*
- 49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq:*
- 50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq:*
- 51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq:*
- 52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq:*
- 53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq:*
- 54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq:*
- 55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq:*
- 56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq:*
- 57: /cgn2_6/ptodata/1/pna/US095_COMB.seq:*
- 58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
- 59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq:*
- 60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq:*
- 61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq:*
- 62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq:*
- 63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq:*
- 64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq:*
- 65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq:*
- 66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq:*
- 67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq:*
- 68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq:*
- 69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
- 70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
- 71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
- 72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq:*
- 73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq:*
- 74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq:*
- 75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq:*
- 76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq:*
- 77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq:*
- 78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq:*
- 79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq:*
- 80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq:*
- 81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq:*
- 82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq:*
- 83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq:*
- 84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq:*
- 85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq:*
- 86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq:*
- 87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq:*
- 88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
- 89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq:*
- 90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq:*
- 91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq:*
- 92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq:*
- 93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq:*
- 94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq:*
- 95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq:*
- 96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq:*
- 97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq:*
- 98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
- 99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
- 100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
- 101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
- 102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
- 103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	542	100.0	542	43	US-09-215-435-78
2	542	100.0	565	76	US-60-096-116-61
3	537.2	99.1	680	1	PCT-US99-17130-73

Sequence 78, Appl
Sequence 61, Appl
Sequence 73, Appl

RESULT 5
US-09-443-634-1
; Sequence 1, Application US/09443634
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: Collipase-Like Protein - 1
; FILE REFERENCE: 99-85
; CURRENT APPLICATION NUMBER: US/09/443,634
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (1)...(354)
US-09-443-634-1

APPLICANT: Chalup, Michael S.
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: BIOCHEMICAL PATHWAY MOLECULES
FILE REFERENCE: PT-0115 P
CURRENT APPLICATION NUMBER: US/60/184,698
CURRENT FILING DATE: 2000-02-24
NUMBER OF SEQ ID NOS: 1132
SOFTWARE: PERL Program
SEQ ID NO 694
LENGTH: 478
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 474641.1.1
FEATURE:
LOCATION: 347-368
OTHER INFORMATION: a, t, c, g, or other
US-60-184-698-694

Query Match 79.2%; Score 429; DB 55; Length 523;
Best Local Similarity 100.0%; Pred. No. 5e-66;
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 72.7%; Score 394.2; DB 96; Length 478;
Best Local Similarity 94.1%; Pred. No. 6e-60;
Matches 396; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 114 ggaactcaaggagcttgcacccggaaccaggagactgagactggtgctgccaactgac 173
DB 90 ggaactcaaggagcttgcacccggaaccaggagactgagactggtgctgccaactgac 149
QY 174 tccagacaattgcagtcgactgcgcgagaaagggtccgagggcagctgtgtcaaac 233
DB 150 tccagacaattgcagtcgactgcgcgagaaagggtccgagggcagctgtgtcaaac 209
QY 234 gcaggtgttcttggccaatatagagcgtgtccctgcctgcggaaacctgactgtatata 293
DB 210 gcaggtgttcttggccaatatagagcgtgtccctgcctgcggaaacctgactgtatata 269
QY 294 ttcaagaataagaaatggcttagcactgcctatggcgtgttcagaaattggaaggca 353
DB 270 ttcaagaataagaaatggcttagcactgcctatggcgtgttcagaaattggaaggca 329
QY 354 gaagtggcctaagaaatgttcttctagtcctcctctcttcttcttctcctcctcctc 413
DB 330 gaagtggcctaagaaatgttcttctagtcctcctcctcctcctcctcctcctcctc 389
QY 414 cactgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 473
DB 390 cactgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 449
QY 474 tgagtggaggaagtggggagtgattgaaataaagagcttttcaatgaaaaaataaaaa 533
DB 450 tgagtggaggaagtggggagtgattgaaataaagagcttttcaatgaaaaaataaaaa 509
QY 534 aaaaaaaa 542
DB 510 aaaaaaaa 518

RESULT 6
US-60-184-698-694
; Sequence 694, Application US/60184698
; GENERAL INFORMATION:
; APPLICANT: Hodgson, David M.
; APPLICANT: Lincoln, Stephen E.
; APPLICANT: Russo, Frank D.
; APPLICANT: Spiro, Peter A.
; APPLICANT: Banville, Steve C.
; APPLICANT: Bratcher, Shawn R.
; APPLICANT: Dufour, Gerard E.
; APPLICANT: Cohen, Howard J.
; APPLICANT: Rosen, Bruce
; APPLICANT: Shah, Purvi
; APPLICANT: Jones, Anissa L.
; APPLICANT: Yu, Jimmy Y.
; APPLICANT: Greenawalt, Lila B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Roseberry, Ann M.
; APPLICANT: Wright, Rachel J.
; APPLICANT: Chen, Wensheng
; APPLICANT: Liu, Tommy
; APPLICANT: Yap, Pierre E.
; APPLICANT: Stockdreher, Theresa K.

QY 112 ttggagctcaaggagcttgcacccggaaccaggagactgagactggtgctgccaactgac 171
DB 58 ttggagctcaaggagcttgcacccggaaccaggagactgagactggtgctgccaactgac 117
QY 172 gctccagacaattgcagtcgactgcgcgagaaagggtccgagggcagctgtgtcaaac 231
DB 118 gctccagacaattgcagtcgactgcgcgagaaagggtccgagggcagctgtgtcaaac 177
QY 232 acgcaggtgttcttggccaatatagagcgtgtccctgcctgcggaaacctgactgtatata 291
DB 178 acgcaggtgttcttggccaatatagagcgtgtccctgcctgcggaaacctgactgtatata 237
QY 292 tatcaagaataagaaatggcttagcactgcctatggcgtgttcagaaattggaagg 351
DB 238 tatcaagaataagaaatggcttagcactgcctatggcgtgttcagaaattggaagg 297
QY 352 cagaagtggcctaagaaatgttcttctagtcctcctcctcctcctcctcctcctcctc 411
DB 298 cagaagtggcctaagaaatgttcttctagtcctcctcctcctcctcctcctcctcctc 357
QY 412 tccacctgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 471
DB 358 nnnnnnnnnntcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 417
QY 472 catgagtggaggaagtggggagtgattgaaataaagagcttttcaatgaaaaaataaaaa 531
DB 418 catgagtggaggaagtggggagtgattgaaataaagagcttttcaatgaaaaaataaaaa 477
QY 532 a 532
DB 478 a 478

RESULT 7
US-60-184-698-48
; Sequence 48, Application US/60184698
; GENERAL INFORMATION:
; APPLICANT: Hodgson, David M.
; APPLICANT: Lincoln, Stephen E.
; APPLICANT: Russo, Frank D.
; APPLICANT: Spiro, Peter A.
; APPLICANT: Banville, Steve C.
; APPLICANT: Bratcher, Shawn R.
; APPLICANT: Dufour, Gerard E.
; APPLICANT: Cohen, Howard J.
; APPLICANT: Rosen, Bruce
; APPLICANT: Shah, Purvi
; APPLICANT: Jones, Anissa L.
; APPLICANT: Yu, Jimmy Y.
; APPLICANT: Greenawalt, Lila B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Roseberry, Ann M.
; APPLICANT: Wright, Rachel J.
; APPLICANT: Chen, Wensheng
; APPLICANT: Liu, Tommy
; APPLICANT: Yap, Pierre E.
; APPLICANT: Stockdreher, Theresa K.

```

; APPLICANT: Chalup, Michael S.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: BIOCHEMICAL PATHWAY MOLECULES
; FILE REFERENCE: PT-0115 P
; CURRENT APPLICATION NUMBER: US/60/184,698
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 1132
; SOFTWARE: PERL Program
; SEQ ID NO 48
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 474641.1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 347-368
; OTHER INFORMATION: a, t, c, g, or other
; US-60-184-698-48

```

Query Match	70.6%	Score 382.4;	DB 96;	Length 463;
Best Local Similarity	94.3%	Pred. No. 6.8e-58;		
Matches 383. Conservative	0;	Mismatches 23;	Indels 0;	Gaps 0;

Qy	112	ttggagctcaagagtgatttgcattcccggaaccaggactgcgagactggtctgctgccaacgt	171
Db	58	tgggagctcaagagtgatttgcattccggaaaccaggactgcgagactggtctgctgccaacgt	117
Qy	172	gctccagacaatttgcagctcgactgcgcggagagaagggtccgagggcagcttggtgcaaa	231
Db	118	gctccagacaatttgcagctcgactgcgcggagagaagggtccgagggcagcttggtgcaaa	177
Qy	232	acgcagctgtcttttggccaatacagacgctgtccctgcttcggaacactgacttgtata	291
Db	178	acgcaggtgtctcttggccaatacagagcgtgtccctgcttcggaacactgacttgtata	237
Qy	292	tattcaagaatgagaatggcttagcatcgctatggccgttgtcagaaaaattggaagg	351
Db	238	tattcaagaatgagaatggcttagcatcgctatggccgttgtcagaaaaattggaagg	297
Qy	352	cagaagttggctaagaaaaattcttcttagtgcctctctcttcttctgctgcctcctcc	411
Db	298	cagaagttggctaagaaaaattcttcttagtgcctctctcttcttctgctgnnnnnnnn	357
Qy	412	tccacctgctctcctaccagagctctgtgttcaacctgttcccccagagcctccac	471
Db	358	nnnnnnnnnnntccccctaccagagctctgtgttcaacctgttcccccagagcctccac	417
Qy	472	catgagtggagggaagtggggagtgattgaaataaagagctttttc	517
Db	418	catgaatcacagaagatggcagctgattgaaataaagagctttttc	463

RESULT 8
US-08-905-279-44
; Sequence 44, Application US/08905279
; GENERAL INFORMATION:
; APPLICANT: Genset SA
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; TITLE OF INVENTION: EXPRESSED IN TESTIS AND OTHER TISSUES
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible

```

1 OPERATING SYSTEM: Win95
2
3 SOFTWARE: Word
4
5 CURRENT APPLICATION DATA:
6
7 APPLICATION NUMBER: US/08/905,279
8
9 FILING DATE:
10
11 CLASSIFICATION: 536
12
13 ATTORNEY/AGENT INFORMATION:
14
15 NAME: Israelisen, Ned A.
16
17 REGISTRATION NUMBER: 29,655
18
19 REFERENCE/DOCKET NUMBER: GENSET.011A
20
21 TELECOMMUNICATION INFORMATION:
22
23 TELEPHONE: (619) 235-8550
24
25 TELEFAX: (619) 235-0176
26
27 INFORMATION FOR SEQ ID NO: 44:
28
29 SEQUENCE CHARACTERISTICS:
30
31 LENGTH: 421 base pairs
32
33 TYPE: NUCLEIC ACID
34
35 STRANDEDNESS: DOUBLE
36
37 TOPOLOGY: LINEAR
38
39 MOLECULE TYPE: CDNA
40
41 ORIGINAL SOURCE:
42
43 ORGANISM: Homo Sapiens
44
45 TISSUE TYPE: Testis
46
47 FEATURE:
48
49 NAME/KEY: sig_peptide
50
51 LOCATION: 62..130
52
53 IDENTIFICATION METHOD: Von Heijne ma
54
55 OTHER INFORMATION: score 9.8
56
57 OTHER INFORMATION: seq FLLFFFLFLTRG
58
59 PS-08-905-279-44
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

Query Match	66.8%;	Score	361.8;	DB	26;	Length	421;
Best Local Similarity	97.6%;	Pred. No.	2.7e-54;				
Matches	360;	Conservative	6;	Mismatches	3;	Indels	0;
Gaps							

Qy	1	cacgacgtgtggccatgatctaccacaaagtgtgctgtcttctctcttcttc 60
Db	47	CACGACCGTGTGGCCATATGCTACCCCAATGGCTGTGCTGCTTCTCTCTTC 106
Qy	61	ttctcttctctctcacgagggtcatttctccacaaaaataaaccttttgagctc 120
Db	107	TTTCTTCTTCTCCTCACCGGGGCTCACTTCTCCACAAAAATACACCTTTTGGAGTC 166
Qy	121	aaggagtcttgcatctccgaacacgagactgcagactgctgtgccacgtgtccagac 180
Db	167	AAGGAGKSKKGATSGGGAAACAGAGCTCGAGACTGGCTGTGCCAGCTGCTCCAGAC 226
Qy	181	aattgcgagtcgactgcgcgcgagaaggggtccgaggggcagtcctgtgtccaaacgcaggtg 240
Db	227	TAATTCGGAGTCGCACTGCGCGGAGAAAGGGGTCCGAGGGGCAGTCTGTGTCAAACGCCAGGTG 286
Qy	241	ttctttggccaataagagcgtgtccctgtcctgcggaaacctgactgtatatattcaaa 300
Db	287	TTCTTTGGCCAAATAGACGCTGTCCCTGCCTCGGGAACCTGACTTGTATATATTCAAAG 346
Qy	301	aatgagaaaatgcttagcatcgcttatggccgttgttcagaaaatgtgaagggcagaagttg 360
Db	347	AATGAGAAATGGCTTAGCATCGCTATGCGCCGTGTGTCAGAAAATTGGAAGGCAGAAAGTTG 406
Qy	361	gctaagaaa 369
Db	407	GCTAGRAAA 415

RESULT 9
US-08-196-481-520
; Sequence 520, Application US/08196481
; GENERAL INFORMATION:
; APPLICANT: Haseltine, William
; APPLICANT: Rosen, Craig
; APPLICANT: Ruben, Steve
; APPLICANT: Dillon, Patrick


```

Query Match          41.5%; Score 224.8; DB 97; Length 514;
Best Local Similarity 98.8%; Pred. No. 2.3e-30;
Matches 237; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY      1  cagagactgtggccatgatgctaccccaatggctgctgctgttcctctcttcttc 60
DB      84  cagagactgtggccatgatgctaccccaatggctgctgctgttcctctcttcttc 143
QY     161  ttctctctctccacaggggctcactttctccacaaaataaaccttttgagctc 120
DB     144  ttctctctctccacaggggctcactttctccacaaaataaaccttttgagctc 203
QY     121  aaggagtttgatcccggaaccaggaactgagagactggctgctgctccaaactgctccagac 180
DB     204  aaggagtttgatcccggaaccaggaactgagagactggctgctgctccaaactgctccagac 263
QY     181  aattgcagctgcactgcgcgagaaggggtcccgagggcagctctgtgtcaaacgcaggtg 240
DB     264  aattgcagctgcactgc-cggagaaggggtcccgagggcagctctgtgtcaaacgcaggtg 322

RESULT 14
US-08-992-332-1456
; Sequence 1456, Application US/08992332
; GENERAL INFORMATION:
; APPLICANT: Gooding, Douglas H.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Lane, John C.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Akerblom, Ingrid E.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Snable, James L.
; APPLICANT: Naughton, Rebecca E.
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
; TITLE OF INVENTION: HUMAN PITUITARY
; NUMBER OF SEQUENCES: 2868
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,332
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/033,364
; FILING DATE: DECEMBER 17, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/728,362
; FILING DATE: OCTOBER 9, 1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PD-0072-1 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1456:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 base pairs

```

Wed May 24 09:19:59 2000

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: 2084384H1
; US-60-033-364-14

Query Match 37.5%; Score 203.4; DB 64; Length 263;
Best Local Similarity 99.0%; Pred. No. 1.2e-26;
Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	112	ttggagctcaaggagctcttgcacccggaaccaggactgcgagactggctgctgccaacgt	171
Db	58	TGGGAGCTCAAGGAGTCTTGCATCCGGAACCAAGGACTGGCTGCTGCCAACGT	117
Qy	172	gtccagacaattgcgagtcgacactgcgaggaagggtccgagggcagctctgtgtcaa	231
Db	118	GCTCCAGACAAATTCGAGTCCACTGCCGAGAGAGGGGTCCGAGGGCAGCTGTGTCAA	177
Qy	232	acgcaggtgtcttggccaatatagagcgtgtccctgcctgcggaacctgacttgata	291
Db	178	ACGCAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGGGGAACCTGACTTGTA	237
Qy	292	tattcaagaatgagaaatggcttag	317
Db	238	TATTCAAGAATGAGAAATGGCTTAG	263

Search completed: May 24, 2000, 01:01:08
Job time: 14240 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:17:06 ; Search time 6115.43 Seconds
(without alignments)
-142.051 Million cell updates/sec

Title: US-09-215-435-113
Perfect score: 893
Sequence: 1 gccatgacgcacgtcac.....aaaaaaaaaaaaaaaa 893

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- GenEmbl:*
1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_inl:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*

- 45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	797.2	89.3	4035	40	AF160973	AF160973 Homo sapi
2	797.2	89.3	4054	9	AB032994	AB032994 Homo sapi
3	797.2	89.3	4126	11	HSM801523	AL136549 Homo sapi
4	683.6	76.6	1117	12	AF162472	AF162472 Mus muscu
5	571.8	64.0	4379	9	HUMHA1025A	D38549 Human mRNA
6	554.2	62.1	4112	12	AF072697	AF072697 Mus muscu
7	376	42.1	51479	44	AC019911	AC019911 Drosophil
8	216	24.2	87184	41	AC009185	AC009185 Homo sapi
9	216	24.2	111321	52	AC016571	AC016571 Homo sapi
10	187.6	21.0	111321	52	AC016571	AC016571 Homo sapi
11	96.6	10.8	160386	43	AC016446	AC016446 Homo sapi
12	83.6	9.4	35382	34	CELF56A11	AF038619 Caenorhab
13	78	8.7	160386	43	AC016446	AC016446 Homo sapi
14	50.6	5.7	10660	2	AE001963	AE001963 Deinococc
15	44.4	5.0	4038	1	MXALON	D12923 Myxococcus
16	44.4	5.0	10847	2	AF127082	AF127082 Myxococcu
17	41.2	4.6	43147	1	SC4A10	AL109663 Streptomy
18	41	4.6	4582	7	CRU16725	U16725 Chlamydomon
19	40.6	4.5	974	1	S77084	S77084 catechol 2,
20	40.6	4.5	96964	33	AC004906	AC004906 Homo sapi
21	40.2	4.5	15742	2	AF013216	AF013216 Myxococcu
22	40.2	4.5	17741	11	HS402G11	AL022328 Human DNA
23	40	4.5	59762	7	AB023032	AB023032 Arabidops
24	40	4.5	74632	33	HSJ132F21	AL079335 Homo sapi
25	39.8	4.5	3800	1	AVIH2AA	M33152 A.vinelandi
26	39.8	4.5	13914	1	AVIHOXHY	L23970 Azotobacter
27	39.6	4.4	1338	7	CRU16825	U16825 Chlamydomon
28	39.6	4.4	2746	7	NCACU9	X56672 N.crassa ma
29	39.4	4.4	1950	1	AVIGLNA	M57275 A.vinelandi
30	39.4	4.4	3692	9	HSGLYPIC	X54232 Human mRNA
31	39.4	4.4	7574	1	AEODHABL	X91877 A.eutrophus
32	39	4.4	1976	12	RN005395	AJ005395 Rattus no
33	39	4.4	2217	12	RNP01C	X70369 R.norvegicu
34	39	4.4	284673	41	AC006028	AC006028 Homo sapi
35	38.8	4.3	3574	12	AF185613	AF185613 Mus muscu
36	38.6	4.3	10629	2	AE002052	AE002052 Deinococc
37	38.6	4.3	182208	52	AC017084	AC017084 Homo sapi
38	38.4	4.3	717	8	CRU61364	U61364 Chlamydomon
39	38.4	4.3	23204	8	CRE243806	AJ243806 Chlamydom
40	38.4	4.3	154746	16	HSV2HG52	Z86099 Herpes simp
41	38.2	4.3	2267	8	AF149803	AF149803 Zea mays
42	38.2	4.3	2502	5	AR023963	AR023963 Sequence
43	38.2	4.3	2502	5	AR054700	AR054700 Sequence
44	38.2	4.3	2502	5	AR061696	AR061696 Sequence
45	38.2	4.3	2502	5	AR061937	AR061937 Sequence

ALIGNMENTS

```

RESULT 1
AF160973 4035 bp mRNA PRI 28-JUL-1999
LOCUS Homo sapiens p53 inducible protein (PIR121) mRNA, complete cds.
DEFINITION
ACCESSION AF160973
VERSION AF160973.1 GI:5616319
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4035)
AUTHORS Saller,E., Tom,E., Brunori,M., Otter,M., Estreichner,A., Mack,D. and
1990,R.
TITLE Increased apoptosis induction by 121F mutant p53
JOURNAL EMBO J. (1999) In press
REFERENCE 2 (bases 1 to 4035)
AUTHORS Saller,E. and Iggo,R.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1999) Oncogene Group, Swiss Institute for
Experimental Cancer Research, Ch. des Boveresses 155, Epalinges
CH-1066, Switzerland
FEATURES
Location/Qualifiers
1..4035
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q34"
1..4035
/gene="PIR121"
49..3810
/gene="PIR121"
/codon_start=1
/product="p53 inducible protein"
/protein_id="AAD45723.1"
/db_xref="GI:5616320"
/translation="MTHTLTEDALSNVLLLELPDQPCIEPPSPISIMYQANFOT
NFERDPAFTGIARYIQEATVSHSNMELPEGHVAVMLYTWRSCEAIQVCKNEOP
NRVEIYKTVLEPEVTKLMKMFYFKAIERCSSEVKRLCHAEKRFQVSEAYLLT
LGKTFINMFADVELLKNKSCVKNDSAYKRAAQFLKRMADPQSQESQNSLMFLANH
RITQLHQHLEIPCYEELADIVNICVYKAYEENKMLTPEKHMLLMGFGFLYMDG
NSNITYKIDNCKEOMVQIRDDHIFISELARYSNSEVVTGSLDSQESDEYRELFDL
TQSSIPDYNICQVQVPLFDMQIELARYIKTSYAHSEENSKWTC
ALRGQLLSKWSAHMVEYSWKLVPDCKFCNCPGTAERYERATRYNTSEKFAF
VEVTAMKGLVLMGRSESVNQAIRNTIYAALQDFQVTLREPLQAVRKKNVLS
VLQARKTICDWEGRPPNDPLURGEKDPKGGEDIKVPRAVGPSTQLYWRVMLE
SLIADKSGSKTTLRSSLDGPVIAIEDFHQSQFFETHLLNISEALQCCDLSQLWFR
FFLELTGRRTOFTIEMSPWILTDHILETKPSMMEYVLYPLDLYNDSAYYATKFK
KOFLYDETEAEVNLFCDFQVYKADQIFAYYKAMAGSVLLDKRFAECKNKGVIIPY
PSNRYETLLKORHQLLGRSIDLNLITQRIISAAMYSKLDQAISEFSEDLTSIVELE
WLEINRLTHRLCKHMTLDSDFAMFEANHNVSAPYGRITLHVFWELNDFLPNYCY
NGSTNRFRVTAIPTQEPQRKDPANQVQYILYSGKPLNIAYSHIYSSYRNFVGPHEK
TICRLGQVGTAVNVEELLKIVKSLLOQTILQYVKTLEIEMPKICRLPRHEYSGPIL
EFFHQLKDIIEYAEKTDVQSLREVGNAIFCLLIEQALSQBEVCDLLHAAPFQNI
LPRVYIKGEERLEVMKRLKAYAPLHLVPLIERLGTPOQIATIAREGDILLTKERLCOG
LSMPEVITLRSYLODPPIWRGPPPTNGVHDECFEHLRLWSAQVFCIPVGTNEF
TAQCQSGGLNAGCSIIVLGQORREDFECYHLLKVRQDCKDEILKNVPLKKA
DRIRKYLINNEVFALNKKYKSVETSDSVIEHVRCRQFPPIHQSLATTC"
BASE COUNT 1020 a 1110 c 1037 g 868 t
ORIGIN
Query Match 89.3%; Score 797.2; DB 40; Length 4035;
Best Local Similarity 99.6%; Pred. No. 1.1e-202;
Matches 799; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 gccatgatcacgcagtcacccctggagatgccctgtccacagctggacctgttgagag 60
|||||
DB 46 GCCATGACCACGACGTCACCCCTGGAGATGCCCTGTCCACAGTGGACCTGCTTGAAG 105
|||||
QY 61 cttccctcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 120
|||||

```


dh10b; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="amygdala"
/map="5q34"
139..3900
/gene="DKF2p761i12121"
139..3900
/note="Homo sapiens p53 inducible protein (Pir121) mRNA,
complete cds."
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB66484.1"
/db_xref="GI:6807649"
/translation="MTHTVLEDAISNVLDLEELPLDPOQPCIEPPSSIMYQANFDT
NFEARNVTVGIARYIEQATVHSSNMNLEEGHYAVMLYTWRSKCSRAIPQVKNQEP
NRVEYKTEVLEPEVTKLMKFWYFQKAIERFCSVKKRLCHAEKRDVSEAYLLT
LGFKNMFAVLDELKMKCSVKNDHSAYKRAAQLKMAQPOSIOESQNSMFLAHN
RITCLOHOLEVIPGYEELLADIVNICVDIYENKMLTPSEKMLLKVMGFGLYLMDG
NVSNIYKLDKAKKRLSKIDKFFKQLOVPLFGDMQIELARYIETSAHYEENKSKWTC
TQSSIPQYNICEQMVQIRDDHIFRISELARYSNSEVTVSGSLDSQKSDDEYRELEFDL
ALRGQLLSKSAHMEVYSKLVHPDKFCNKDCPCGTAEYERATRYNTSEKFAF
VEVAMIKGLOVLMGRMSEVNOAIRNTIYAALODFAQVTLREPLRAQVRKKNVLIS
VLOAKRTICDMEGRSPNDPCIRGKDPKGFDIKVPRAVGPSTQLYMYMTMLE
SLIADKSKKTKTLASSLDGPVYLAIEDPHKQSFETHLLNLSALQCCDLSQLWRE
FFLEUTMGRRIOQPIEMSPWILTDHILETRKPSMMEYVLYPLDLYNDSAYALTKPK
KQFLYDEIEAEVNLGDFQVYKQADQIFAYIKAMAGSVLLDOKRFRACBCKNVTIIP
PSNRYETLLKQHVQLLRSLDNLRLITQRIASAMYSLOQATSRFSEDLTSIVLE
WLEENRTHRLCKHMTLDSFDAMREANHNVSAPYGRITLHFVFNELNDFLPNICY
NGSTNRVFTAPFTQEPQORDKANVOPIYLYGSKPLNIAYSHIYSSYRNFVGPPEK
TICRLLYGQIYAVMEELKIVKSLQCTILQYVKTLEVMKICRLPRHIEYGGPIL
EFFHQLADIEYAEKLDVQSLEVGNAILFCGLLEALSOREVCDLLHAAPFQNT
LPRVYIKERLEVRMKLEAKYAPLHLVPLIERLGTQPOIARREGDLLTKERLGG
LSMFEVILTRINSYLODPPIWGRPPPTNGMHVDECFEVRWSAQWYFICIPVTNEF
TAEQCFGLGLWAGCSIIYLLGQRRDLDFCYHLLKVRQDGDLEIKNVLPKKMA
DRIRYQILNNEVAILNKYMKSVETDSTVEHVRCEFPPIHQSLATTC"
4102
polvA_site
BASE COUNT 1028 a 1143 c 1081 g 874 t
ORIGIN
Query Match 89.3%; Score 797.2; DB 11; Length 4126;
Best Local Similarity 99.6%; Pred. No. 1.1e-202;
Matches 799; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 gccatgacgcacgtcaccttggaagatgccctgtccaaagtggaacctgttgaagaag 60
Db 136 gccatgacgcacgtcaccttggaagatgccctgtccaaagtggaacctgttgaagaag 195
Qy 61 ctccccctcccgacagcagcgcacgtcgcacgtccacccctccctccatcatgtaccag 120
Db 196 CTTCCTCCCTCCCGACGACGACCATGCATCGACGCTCCACCTTCCTCCATCATGTACCAG 255
Qy 121 gctaaacttgacacaaacttgaggacaggaatgcatctgttcacgggacattgcaaggtac 180
Db 256 GCTAACTTTGACACAAACTTTGAGGACAGGAATGCATTTGTACACGGGCGATTGCAAGGTAC 315
Qy 181 attgagaggctacagtcceactcacatgaatgaatgctgaggaagacatgagtat 240
Db 316 ATTGAGCAGGCTACAGTCCACTCCAGCATGAATGATGCTGTGAGGACGACATGAGTAT 375
Qy 241 gcggtcatgctgtacacactggcgagcgtgttcccggggcaattccccagggtgaatgcaac 300
Db 376 GCGGTGCTGTACACTGCGCGAGCTGTTCGCGGCGCATTCCTCCAGGTGAAATGCAAC 435
Qy 301 gagcagcccaaccagtagagatctatgagaagacagtagagtgctgagcagcgaggtc 360
Db 436 GAGCAGCCCAACCCAGTAGAGATCTATGAGAAGACAGTAGAGGTGCTGTGAGCGGAGGTC 495
Qy 361 accaagctcatgaagttcatgtattttcagcgcaaggccatcagcggttcttcgacgcag 420
Db 496 ACCAAGCTCATGAAGTTCATGTATTTTCAGCGCAAGGCCATCGAGCGGTCTTCGACGGAG 555

Qy 421 gtgaagcgcgtgtgccatgccagcgcaggaagagcatttctctctgagcctacacctcctg 480
Db 556 GTGAAGCGCGTGTGCCATGTCGCGAGCGCAGAGGAGCTTTGTCTCTGAGCGCTACCTCTCG 615
Qy 481 acccttggcaagtctcaacatgtttgtctctgagtagctgagcgaagacatgaagtcg 540
Db 616 ACCCTTGGCAAGTTCATCAACATGTTTGTCTCTGAGTGAAGTGAAGTGAAGTGC 675
Qy 541 agcgtcaagaatgaccactccgctcacaagaggagcagcagcttctcgcgaagatggca 600
Db 676 AGCGTCAAGAATGACCACTCTGCCTACAAAGAGGCGACAGTTCCTCGGGAAGATGGCA 735
Qy 601 gatccccagctctaccagagtcgcagaaaccttccatgttctctggcccaaccacacag 660
Db 736 GATCCCCAGCTCTATCCAGGAGTGCAGAACCTTTCCATGTCTTCTGCGCAACACAGG 795
Qy 661 atcacccagtgctccaccagcaactgaagtgaatccccaggctatgagagctgctgct 720
Db 796 ATCACCCAGTGTCTCCACCAGCAACTTGAAGTGTATCCAGGCTATGAGGAGCTGCTGGCT 855
Qy 721 gacattgtcaacatctgtgtgagttactacgagaaagatgtacctgactccacagtgag 780
Db 856 GACATTGTCAACATCTGTGTGGATTTACTACGAGACACAGATGTACTGACTCCACAGTGAG 915
Qy 781 aacatatgtctctcaaggtaa 802
Db 916 AACATATGCTCCTCAAGGTGA 937
RESULT 4
AF162472 1117 bp mRNA ROD 28-JUL-1999
DEFINITION Mus musculus inducible protein (Pir121) mRNA, partial cds.
ACCESSION AF162472
VERSION AF162472.1 GI:5616489
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1117)
AUTHORS Saller,E., Tom,E., Brunori,M., Otter,M., Estreicher,A., Mack,D. and Iggo,R.
TITLE Increased apoptosis induction by 121F mutant p53
JOURNAL EMBO J. (1999) In press
REFERENCE 2 (bases 1 to 1117)
AUTHORS Saller,E. and Iggo,R.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-1999) Oncogene Group, Swiss Institute for Experimental Cancer Research, Ch. des Boveresses 155, Epalinges CH-1066, Switzerland
FEATURES
source 1..1117
/organism="Mus musculus"
/db_xref="taxon:10090"
1..1117
/gene="Pir121"
16..>1117
/gene="Pir121"
/note="Pir121"
/codon_start=1
/product="inducible protein"
/protein_id="MAD45803.1"
/db_xref="GI:5616490"
/translation="MTHTVLEDAISNVLDLEELPLDPOQPCIEPPSSIMYQANFDT
NFEARNVTVGIARYIEQATVHSSNMNLEEGHYAVMLYTWRSKCSRAIPQVKNQEP
NRVEYKTEVLEPEVTKLMKFWYFQKAIERFCSVKKRLCHAEKRDVSEAYLLT
LGFKNMFAVLDELKMKCSVKNDHSAYKRAAQLKMAQPOSIOESQNSMFLAHN
RITCLOHOLEVIPGYEELLADIVNICVDIYENKMLTPSEKMLLKVMGFGLYLMDG
NVSNIYKLDKAKKRLSKIDKFFKQLOVPLFGDMQIELARYIETSAHYEENKSKWTC
TQSSIPQYNICEQMVQIRDDHIFRISELARYSNSEVTVSGSLDSQKSDDEYRELEFDL
ALRGQLLSKSAHMEVYSKLVHPDKFCNKDCPCGTAEYERATRYNTSEKFAF
VEVAMIKGLOVLMGRMSEVNOAIRNTIYAALODFAQVTLREPLRAQVRKKNVLIS
VLOAKRTICDMEGRSPNDPCIRGKDPKGFDIKVPRAVGPSTQLYMYMTMLE
SLIADKSKKTKTLASSLDGPVYLAIEDPHKQSFETHLLNLSALQCCDLSQLWRE
FFLEUTMGRRIOQPIEMSPWILTDHILETRKPSMMEYVLYPLDLYNDSAYALTKPK
KQFLYDEIEAEVNLGDFQVYKQADQIFAYIKAMAGSVLLDOKRFRACBCKNVTIIP
PSNRYETLLKQHVQLLRSLDNLRLITQRIASAMYSLOQATSRFSEDLTSIVLE
WLEENRTHRLCKHMTLDSFDAMREANHNVSAPYGRITLHFVFNELNDFLPNICY
NGSTNRVFTAPFTQEPQORDKANVOPIYLYGSKPLNIAYSHIYSSYRNFVGPPEK
TICRLLYGQIYAVMEELKIVKSLQCTILQYVKTLEVMKICRLPRHIEYGGPIL
EFFHQLADIEYAEKLDVQSLEVGNAILFCGLLEALSOREVCDLLHAAPFQNT
LPRVYIKERLEVRMKLEAKYAPLHLVPLIERLGTQPOIARREGDLLTKERLGG
LSMFEVILTRINSYLODPPIWGRPPPTNGMHVDECFEVRWSAQWYFICIPVTNEF
TAEQCFGLGLWAGCSIIYLLGQRRDLDFCYHLLKVRQDGDLEIKNVLPKKMA
DRIRYQILNNEVAILNKYMKSVETDSTVEHVRCEFPPIHQSLATTC"
BASE COUNT 296 a 295 c 295 g 231 t
ORIGIN

Query Match		76.6%;	Score 683.6;	DB 12;	Length 1117;	
Best Local Similarity		90.8%;	Pred. No. 2.3e-172;			
Matches 728;		Conservative	0;	Mismatches 74;	Indels	0; Gaps
QY	1	gcacatgacacgcacgcacacccctggaagatgcccctgtgtccacgctgagacctgttggaagag	60			
Db	13	GCATGACCAACCCACGCTCATTTTGGAAAGATGCCCTGTCTCAACGCTGACCTCTCTGGAGAA	72			
QY	61	cttccctcccgaccagcagcagccatgcatcagcctcagcctccaccttctccatcatgtaccag	120			
Db	73	CTGCCCTCCAGACAGCAGCAGCATGATFCGAGCCCCGCCCATCTCCATCATGTACCAG	132			
QY	121	gctaaactgtgacacaaactttgagacaggaatgcatgttgcaagggcattgcaaggtac	180			
Db	133	GCTAACTTTGACACGAACCTTCGAGGACAGGAATGATTTGTACCGGGCATTTGCAAGGTAC	192			
QY	181	attgagcaggctacagctccactccagcatgaatgagatgctggagggaaggacatgagtat	240			
Db	193	ATTGACGAGCAACCGTCCACTCCAGCATGAATGAATGCTTGGAGAGGGCCACGACTAT	252			
QY	241	gcggtcatctgtacacctggcagcagctgttcccgggccattccccagtgaaatgaac	300			
Db	253	GCAGTCATGCTGTACACCTGGCGGACGCTGCTCCCGGGGCATCCCGCAGGTGAAGTCAAT	312			
QY	301	gagcagcccaaccagtagatgatatgagagacagtagaggtgctggagccgaggtc	360			
Db	313	GAGCAGCCCAATCGGTCGAGATCTACGAGAGACGGTGGAGGTCTGGAGCCAGAGTGC	372			
QY	361	accagtcctgaagtctatgtattttcagcgcaagggccatcgagcgtttctgcagcgag	420			
Db	373	ACCAAGCTCATGAAGTTCATGTACTTTCAGCGCAAGGCCATCGAGCGCTTCTGTAGCGAG	432			
QY	421	gtaaacgctgtgccaatccagcagcaggaagactttgtctctgagccctactcctg	480			
Db	433	GTGAAGCGGCTGTGTCACCGCTGAGCGCAGGAAGGACTTTGTCTGTGAGGGCTTACTGTGCT	492			
QY	481	acccttggcaagtctcaacatgtttgctgtcctggatgagctaaagaacatgaatgc	540			
Db	493	ACGCTGGCAAGTTCATCAACATGTTCCGGGCTTTGGAGGAGCTAAAGAACATGAAGTGC	552			
QY	541	agcgtcaagaatgaccactccgctacagaagggcagcacagttcctcggaagatggca	600			
Db	553	AGTGTCAAGAATGACCACCTCTGCTCAAGAGAGCAGCACAGTTCTCTAAGGAAGATGGCG	612			
QY	601	gatccccagttatccagagtcgagacaccttccatgttcttcctgccaacacacagg	660			
Db	613	GACCCCGATCCATCCAGGAATCTCAGAACCTGTCCATGTTCTTCCCTGGCCCAATCAACACG	672			
QY	661	atcaccagtgctccaccagcaactgaagtgaatgccagggctatgagagctgctggct	720			
Db	673	ATCACCAGTGTCTTACCAGCACTGGGAAGTGAATCCCGGGCTATGAGAGACTGCTGGCA	732			
QY	721	gacattgtcaacatctgtgtgattactacgagaagaagatgtaactgactccccagtgag	780			
Db	733	GACATTGTCAACATCTGTGTGATTACTATGAGATAAGATGTACCTGACGCCCCAGTGAA	792			
QY	781	aaacatatgctctcaaggtaa	802			
Db	793	AAGCACATGTTGCTCAAGGTGA	814			
RESULT 5						
HUMH1025A						
LOCUS	HUMH1025A	4379 bp	mRNA	PRI	10-JUL-1997	
DEFINITION	Human mRNA for KIAA0068 gene, partial cds.					
ACCESSION	D38549					
VERSION	D38549.1 GI:559702					
KEYWORDS	KIAA0068					
SOURCE	Homo sapiens male myeloblast cell-line KG-1 cDNA to mRNA.					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;					

Db	116	CCCTGCCGACCAGCAGCCCTGTCATCGAGCCCGCCGCGCATCTCGCTGCTCTACAGCCA	175
Qy	124	aactttgacacaaatttgaggacaggaatgcatattgtcagcgcatcgattgaagtcacatt	183
Db	176	AATTTCAACACATAACTTTGAACACAGAATAATGATTGTACTGGCATCGCAAGATACATT	235
Qy	184	gagcaggtcacagtcacactccagtcgaatgagatgctggagggaagacatgtagtcgcg	243
Db	236	GAACAGCCACCGTCCACTCTAGCATGAACGAGATGCTGGAGGAGGGCCAAGAATATGCT	295
Qy	244	gtaatgctgtaacactggcgcagctgttcccgggccattccccaggtgaatgaacagag	303
Db	296	GTCAATGCTGTACACTTGGAGGAGTGCTCTCCGGGGCCATCCACAGAGTGAATGTAAACG	355
Qy	304	cagcccaacccgagtagagatctatgagaagacagtagaggtgctgagccgagagtcacc	363
Db	356	CAGCCTAACAGAGTGAATCTACGAGAAAACCGTGGAGGTTCTGGAGCCTGAGTCCACA	415
Qy	364	aagtcctgaagtctcatgtattttaaagcgaaggccatgagcgggttctgcagcagagtg	423
Db	416	AAACTGATGAATTTCAATGACTTCCAGAGAAATGCCATTGAGCGTTCTGCGGGGAAGTG	475
Qy	424	aagcgtctggccatccgagcgcaggaaggactttgtctgagggcctaccctcctgacc	483
Db	476	AGGCGCTGTGCCATGCCGAGAGGAGGAGGACTTCGTGTCAAGAAGCCTACCTGATCACA	535
Qy	484	cttggcaagtctcaacatgtttgtgctcctggatgagctaaagaacatgaagtgcagc	543
Db	536	CTGGCAAAATTCATCAACATGTTTCGTGTGCTGGAGAGCTGAAGAACATGAAGTGCAGT	595
Qy	544	gtcaagaatgaccactccgcctcaagagggcagcacagtctctgcggaagatggcagat	603
Db	596	GTGAAGAAGCAACCACTCAGCGTACAGAGGGCGCTCAGTTTTTACGTAATATGGCAGAT	655
Qy	604	cccagctctaccaggagtcgcagaaacctttccatgttccctggcccaaccacacaggatcc	663
Db	656	CCACAGTCCATCCAGGAATCGCAGAATCTGTCCATGTCTGTGCCAATCATACAAGATC	715
Qy	664	accagtgctctcaaccagcaacitgaagtatcccgagctatgagagctgctggctgcac	723
Db	716	ACACAGCTCTGTGCAGCAGCAGCTCGAAGTAGTATTCTGGCTACGAAGACTCTGGCAGAT	775
Qy	724	attgtcaacatctgtgtggattactacgagaacaagatgtacctgactccccagtgagaaa	783
Db	776	ATTGTGAATCTGTGTGGTATTACTACGAGAACAGGATGTATTTGACGCCAGTGAGAAA	835
Qy	784	catatgctctccaaggtaa	802
Db	836	CACATGCTTCTCAAAGTCA	854
RESULT	6		
AF072697	AF072697	4112 bp	ROD
LOCUS	Mus musculus SHYC (Shyc)	mRNA	08-JUL-1998
DEFINITION	Mus musculus SHYC (Shyc) mRNA, complete cds.		
ACCESSION	AF072697		
VERSION	AF072697.1	GI:3293550	
KEYWORDS	house mouse.		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 4112)		
AUTHORS	Koester,F., Schinke,B., Niemann,S. and Hermans-Borgmeyer,I.		
TITLE	Identification of Shyc, a novel gene expressed in the murine developing and adult nervous system		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 4112)		
AUTHORS	Koester,F., Schinke,B., Niemann,S. and Hermans-Borgmeyer,I.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-JUN-1998) University of Hamburg, Center for Mole		
FEATURES	Neurobiology, Martinistr. 52, Hamburg D-20246, Germany Location/Qualifiers		

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 7, 1999 this sequence version replaced g1:5705980.
www.jgi.doe.gov.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 14428: contig of 14428 bp in length
* gap of unknown length
* 14429 87184: contig of 72756 bp in length.

FEATURES
source
1. 87184
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CIR-HSPC-248019"
BASE COUNT 25707 a 18905 c 18505 g 24067 t
ORIGIN

Query Match 24.2%; Score 216; DB 41; Length 87184;
Best Local Similarity 100.0%; Pred. No. 5.1e-47; Indels 0; Gaps 0;
Matches 216; Conservative 0; Mismatches 0;
QY 667 cagtgctccaccagcaactgaagtatcccgagctatgagagctgctgctgacatt 726
|||||
Db 9111 CAGTGTCTCCACGACCACTTGAAGTGATCCCGAGCTATGAGGAGCTGCTGGCTGACATT 9052
QY 727 gtoaacatctgtgtgattactacagagaacaagatgtacctgactcccgagtgagaacat 786
|||||
Db 9051 GTCAACATCTGTGTGATTACTACGAGAACAAAGATGTACCTGACTCCCGAGTGAGAACAT 8992
QY 787 atgctctcaagtaaaactccctgagcgccaccatggagctgggcttaccctctc 846
|||||
Db 8991 ATGCTCTCAAGTAAATCCCTCGAGCGCCGCCATCGAGCTGGCTTACCTCTC 8932
QY 847 acctcttcttataaaatccgttttaaaaaacaa 882
|||||
Db 8931 ACCTTCTCTTATTAATAATCCGTTTAAAAAACAA 8896

RESULT 9
AC016571/c
LOCUS
DEFINITION AC016571 111321 bp DNA HTG 27-JAN-2000
Homo sapiens chromosome 5 clone CTB-109A12, WORKING DRAFT SEQUENCE,
39 unordered pieces.
ACCESSION AC016571
VERSION AC016571.2 GI:6778425
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM human.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
TITLE Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 111321)
REFERENCE Direct Submission.
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 111321)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 27, 2000 this sequence version replaced g1:6524135.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics

Consensus quality: 68593 bases at least Q40
Consensus quality: 86380 bases at least Q30
Consensus quality: 92425 bases at least Q20
Estimated insert size: 111321; sum-of-contigs estimation
Estimated insert size: 90000; pulse field gel estimation
Quality coverage: 3.49x in Q20 bases; pulse field gel estimation
Quality coverage: 2.82x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1122: contig of 1122 bp in length
* gap of unknown length
* 1123 2162: contig of 1040 bp in length
* gap of unknown length
* 2163 3309: contig of 1147 bp in length
* gap of unknown length
* 3310 4485: contig of 1176 bp in length
* gap of unknown length
* 4486 5662: contig of 1177 bp in length
* gap of unknown length
* 5663 6685: contig of 1023 bp in length
* gap of unknown length
* 6686 7749: contig of 1064 bp in length
* gap of unknown length
* 7750 8769: contig of 1020 bp in length
* gap of unknown length
* 8770 9815: contig of 1046 bp in length
* gap of unknown length
* 9816 11093: contig of 1278 bp in length
* gap of unknown length
* 11094 13023: contig of 1930 bp in length
* gap of unknown length
* 13024 14059: contig of 1036 bp in length
* gap of unknown length
* 14060 15351: contig of 1292 bp in length
* gap of unknown length
* 15352 16409: contig of 1058 bp in length
* gap of unknown length
* 16410 17589: contig of 1180 bp in length
* gap of unknown length
* 17590 18806: contig of 1217 bp in length
* gap of unknown length
* 18807 19843: contig of 1037 bp in length
* gap of unknown length
* 19844 20847: contig of 1004 bp in length
* gap of unknown length
* 20848 22574: contig of 1727 bp in length
* gap of unknown length
* 22575 24458: contig of 1884 bp in length
* gap of unknown length
* 24459 25574: contig of 1116 bp in length
* gap of unknown length
* 25575 26650: contig of 1076 bp in length
* gap of unknown length
* 26651 28253: contig of 1603 bp in length
* gap of unknown length
* 28254 29913: contig of 1660 bp in length
* gap of unknown length
* 29914 31507: contig of 1594 bp in length
* gap of unknown length
* 31508 32849: contig of 1342 bp in length
* gap of unknown length
* 32850 35538: contig of 2689 bp in length
* gap of unknown length
* 35539 38476: contig of 2938 bp in length
* gap of unknown length
* 38477 41402: contig of 2926 bp in length

```
* 41403 44835: gap of unknown length
* contig of 3433 bp in length
* 44836 48962: gap of unknown length
* contig of 4127 bp in length
* 48963 53805: gap of unknown length
* contig of 4843 bp in length
* 53806 60163: gap of unknown length
* contig of 6358 bp in length
* 60164 66864: gap of unknown length
* contig of 6701 bp in length
* 66865 73707: gap of unknown length
* contig of 6843 bp in length
* 73708 82492: gap of unknown length
* contig of 8785 bp in length
* 82493 88884: gap of unknown length
* contig of 6392 bp in length
* 88885 98392: gap of unknown length
* contig of 9508 bp in length
* 98393 111321: gap of unknown length
* contig of 12929 bp in length.
```

```
FEATURES
  Location/Qualifiers
    1..111321
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="CTB-109A12"
BASE COUNT 30209 a 25928 c 25518 g 29342 t 324 others
ORIGIN
```

```
Query Match      24.2%; Score 216; DB 52; Length 111321;
Best Local Similarity 100.0%; Pred. NO. 5.3e-47;
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 cagtgctccacagcaactgaagtgcacccagggtatgagagtgctggtgcacatt 726
|||||
Db 111021 CAGTGTCTCCACGAGCAACTGAAAGTGATCCCGAGGCTATGAGGAGCTGTGGTGACATT 110962
|||||

QY 727 gtaacatctgtgtgattactacagagaacagatatactgaactccagtgagaacat 786
|||||
Db 110961 GTCACATCTGTGTGGATTACTACGAGACACAGATGACTGACTCCCGATGAGAACAT 110902
|||||

QY 787 atgctctcaagtaaaactccctgagcgccaccatggagcgtgggtaccctctc 846
|||||
Db 110901 ATGCTCCTCAAGTAAATCTCCCTGAGCGCCGACCCATGGAGCCTGGGCTTACCTCTC 110842
|||||

QY 847 accttctttattaaaaatccgttttaaaaaaaca 882
|||||
Db 110841 ACCTTCTTCTATTAAAAATCCGTTTAAAAAACA 110806
|||||
```

```
RESULT 10
AC016571
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTB-109A12, WORKING DRAFT SEQUENCE,
39 unordered pieces.
ACCESSION AC016571
VERSION AC016571.2 GI:6778425
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 111321)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
2 (bases 1 to 111321)
DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL
On Jan 27, 2000 this sequence version replaced gi:6524135.
COMMENT
```

```
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 68593 bases at least Q40
Consensus quality: 86380 bases at least Q30
Consensus quality: 92425 bases at least Q20
Estimated insert size: 111321; sum-of-contigs estimation
Estimated insert size: 90000; pulse field gel estimation
Quality coverage: 3.49x in Q20 bases; pulse field gel estimation
Quality coverage: 2.82x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1122: contig of 1122 bp in length
* gap of unknown length
* 1123 2162: contig of 1040 bp in length
* gap of unknown length
* 2163 3309: contig of 1147 bp in length
* gap of unknown length
* 3310 4485: contig of 1176 bp in length
* gap of unknown length
* 4486 5662: contig of 1177 bp in length
* gap of unknown length
* 5663 6685: contig of 1023 bp in length
* gap of unknown length
* 6686 7749: contig of 1064 bp in length
* gap of unknown length
* 7750 8769: contig of 1020 bp in length
* gap of unknown length
* 8770 9815: contig of 1046 bp in length
* gap of unknown length
* 9816 11093: contig of 1278 bp in length
* gap of unknown length
* 11094 13023: contig of 1930 bp in length
* gap of unknown length
* 13024 14059: contig of 1036 bp in length
* gap of unknown length
* 14060 15351: contig of 1292 bp in length
* gap of unknown length
* 15352 16409: contig of 1058 bp in length
* gap of unknown length
* 16410 17589: contig of 1180 bp in length
* gap of unknown length
* 17590 18806: contig of 1217 bp in length
* gap of unknown length
* 18807 19843: contig of 1037 bp in length
* gap of unknown length
* 19844 20847: contig of 1004 bp in length
* gap of unknown length
* 20848 22574: contig of 1727 bp in length
* gap of unknown length
* 22575 24458: contig of 1884 bp in length
* gap of unknown length
* 24459 25574: contig of 1116 bp in length
* gap of unknown length
* 25575 26650: contig of 1076 bp in length
* gap of unknown length
* 26651 28253: contig of 1603 bp in length
* gap of unknown length
* 28254 29913: contig of 1660 bp in length
* gap of unknown length
* 29914 31507: contig of 1594 bp in length
* gap of unknown length
* 31508 32849: contig of 1342 bp in length
* gap of unknown length
```

Wed May 24 09:19:39 2000

JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

Unpublished
2 (bases 1 to 160386)
Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
Yu,S. and Davis,R.W.
Direct Submission
Submitted (30-NOV-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

----- Genome Center
Center: Stanford DNA Sequencing and Technology Development

Center
Center code: SDBSTD
Web site: <http://sequence-www.stanford.edu/group/human/>
Contact: hum-info@sequence.stanford.edu

----- Project Information
Center project name: 712
Center clone name: RP11-289D12

----- Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-primer; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319
Consensus quality: 126842 bases at least Q40
Consensus quality: 143476 bases at least Q30
Consensus quality: 148432 bases at least Q20

Insert size: 187529; agarose-fp
Quality coverage: 3.2x in Q20 bases; agarose-fp
Quality coverage: 3.8x in Q20 bases; sum-of-contigs.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1065: contig of 1065 bp in length
1066 1115: gap of unknown length
1116 2132: contig of 1017 bp in length
2133 2182: gap of unknown length
2183 3573: contig of 1391 bp in length
3574 3623: gap of unknown length
3624 4704: contig of 1081 bp in length
4705 4754: gap of unknown length
4755 6468: contig of 1713 bp in length
6469 6517: gap of unknown length
6518 7714: contig of 1197 bp in length
7715 7764: gap of unknown length
7765 8891: contig of 1127 bp in length
8892 8941: gap of unknown length
8942 10657: contig of 1716 bp in length
10658 10707: gap of unknown length
10708 12268: contig of 1561 bp in length
12269 12318: gap of unknown length
12319 14128: contig of 1810 bp in length
14129 14179: contig of 1992 bp in length
14180 16220: gap of unknown length
16221 18946: contig of 2726 bp in length
18947 18996: gap of unknown length
18997 21085: contig of 2089 bp in length
21086 21135: gap of unknown length
21136 22394: contig of 1259 bp in length
22395 22445: gap of unknown length
22446 24761: contig of 2317 bp in length
24762 26706: contig of 1895 bp in length
26707 26756: gap of unknown length
26757 29452: contig of 2696 bp in length
29453 29502: gap of unknown length

32850 35538: contig of 2689 bp in length
35539 38476: contig of 2938 bp in length
38477 41402: contig of 2926 bp in length
41403 44835: contig of 3433 bp in length
44836 48962: contig of 4127 bp in length
48963 53805: contig of 4843 bp in length
53806 60163: contig of 6358 bp in length
60164 66864: contig of 6701 bp in length
66865 73707: contig of 6843 bp in length
73708 82492: contig of 8785 bp in length
82493 88884: contig of 6392 bp in length
88885 98392: contig of 9508 bp in length
98393 111321: contig of 12929 bp in length.
Location/Qualifiers
1. .111321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-109A12"
BASE COUNT 30209 a 25928 c 25518 g 29342 t 324 others
ORIGIN

Query Match 21.0%; Score 187.6; DB 52; Length 111321;
Best Local Similarity 97.9%; Pred. No. 2.1e-39;
Matches 190; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 388 cagcgaagccatgcgagcttctgcagcgaggtgaagcggtgtgccatgccagcgc 447
Db 37294 CAGCGAAGGCCATCGAGCGGTTCTGCAGCGAGGTGAAGCGGTGTGCCATGCCGAGCGC 37353
Qy 448 aggaaggaacttctctgagcctacctctgaccccttgcaagttcatcaacatgttt 507
Db 37354 AGGAAGGACTTTGTCTGTGAGGCCCTACTCTCTGACCCCTTGGCAAGTTCATCAACATGTTT 37413
Qy 508 gctgtcctggatgactaaagaacatgaagtgcagcgtcaagaatgaccactccgcttac 567
Db 37414 CTTGTCTCTGGATGAGCTAAAGAACATGAAGTGCAGCGTCAAGAATGACCCTCTGCCTAC 37473
Qy 568 aagaggcagacaca 581
Db 37474 AAGAGTCAAGGCA 37487

RESULT 11
AC016446 150386 bp DNA HTG 30-NOV-1999
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-289D12, *** SEQUENCING IN
PROGRESS ***, 43 unordered pieces.
AC016446
VERSION AC016446.1 GI:6478906
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 160386)
AUTHORS Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
Yu,S. and Davis,R.W.

[illegible]


```
/transl_table=11
/product="hypothetical protein"
/protein_id="AA10711.1"
/db_xref="GI:6458875"
/translation="MVHAPGGGSRVVRDAAGRLRLWPLPAGWLPNVDDGEVRLFSHG
ESIFVTPGSVAVLRGGKQPIPLPAGIPIFNFTPLGRWLGLTGWGTVEDEAGVKV
YQAGAGPAPSGPYVLLPIEGQLTFPGPK"
complement(4797. .5786)
/gene="DR1137"
CDS
complement(4797. .5786)
/gene="DR1137"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA10712.1"
/db_xref="GI:6458876"
/translation="MWDARPMVRFVFPADAPAWAVLNNVNRATPAGLLAEDAR
LAEOVRWVAEEGVVLAHLYFFPLPNFLQDLTVSPDARRRHGHSALWETAL
AETQALGTSIATVRRDDPPSRWAERGFALHAHRFSSRLDLASFOTFPLPALVR
AEAGVTPSNLAGADEATVTRYLDVADRLIETPDLAGHPWEREOVRGMLHKNPR
PDWLVLAGPGEWLTAMVOYRDLAYNELTAVSPQARGRLALPLKLYAIRRAAA
GLSMATNHNHAPMLAIVNRKLGFAPOVGRKWEFLFREGESLAGSVADQPNLLK"
complement(5868. .6632)
/gene="DR1138"
CDS
complement(5868. .6632)
/gene="DR1138"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA10713.1"
/db_xref="GI:6458877"
/translation="MKNELLTALLAGAAQTAQSVTTTTTTPQSTTAPSYASGYT
FRQMDALGALKVELRGVDDRGILNLTVANDAAQQAIAQAAAGLPTGVIAFDG
VPLAGTGVSTTTPAPAGTGTQSSNONSLPTNGAPLAQHTAQSLGSPRTVRAGE
ANTWSFVLSGNTGSTIHLHGACDVRFEVFSASGEIVRPDPTNTICTLQLVVTVDVAPG
EQREVQKIRWEGKNGARQPVPAASTPSPCSGARAP"
complement(6709. .7104)
/gene="DR1139"
CDS
complement(6709. .7104)
/gene="DR1139"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA10714.1"
/db_xref="GI:6458878"
/translation="MSGPATPWGVEWRAGERERFIWRGDVOPYRTEPLPAPVNYG
CLPGLNPADDAEVDVWVGLPEPLAVGTETDLAPSLGLADGDHVKVIFGELKAGLEAL
LAWPPPERGAQVQGAQAQWLAGLSAAE"
complement(7101. .7475)
/gene="DR1140"
CDS
complement(7101. .7475)
/gene="DR1140"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA10715.1"
/db_xref="GI:6458879"
/translation="MKREVLALALASPLIGGCREFVPLIPPATSVTLPTRTITAOEL
KRAGELRLSVTLQGRLEPGYLTWHFVDSGLDGLTDSVFLDAAKTAELTLPAPAGA
TRAVLSFGTVLRQVELIYVRP"
7544. .9040
/gene="DR1141"
CDS
7544. .9040
/gene="DR1141"
/feature="similar to PID:1652891 percent identity: 67.92;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glycosyl hydrolase, family 13"
```

```
/protein_id="AA10708.1"
/db_xref="GI:6458872"
/translation="MSRPPAILLRPMSEPVSAVPLSAPVTEWVADAVFYQIFPPDRFA
RSRVTGLHLPWGSAPTIHTGMGDLMGVAERJGLAGLVNALYFCPVFQSNHR
YHTHDYFOVDPLMGNAALRDLDEAHARGIRVLDGVFNHASRGGFFQNDLHEGEA
YAKRDMFTVGPMSAYDEDRPANYAAWNGIRALPKFNTDHPREFFLLSVAEYWMHF
GIDWRCLDPVNEIDDDAFWQEFRRVKAINPDAYIVGTEWDAWLAGDQDFDAVMNY
HFTRWCLGFFGARTLDHPNERSGTGRVEPLDAAAFVAVRVEVTRMYHPAVVRAOLNL
LDSDHTARFLTAVGSDASAOALAVFOFSYVGAPCLYXGDEIGLPGGVDPDCRRAPFW
DEPOTWHDHTLTILRELTAARHRSALARRGDTFVLHAQGEQVSFAKKHGDGEAYVAFN
CALDAAPVPVPATSGRYRDMLTGAVHDLNESSVILTPARTALLICEG"
9084. .10526
/gene="DR1142"
Query Match 5.7%; Score 50.6; DB 2; Length 10660;
Best Local Similarity 48.8%; Pred. No. 0.0066;
Matches 137; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
QY 342 ggtgctggagcgcggaggtcaccaaactcatgaagttcattgtattttcagcgcaaggccat 401
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1065 GGTGTGCACCTGCTGGCGCGCCGCCGACGAGCGCGGGTGGATTTCACCATGCAGACAT 1006
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 402 cgagcgggtctcagcagaggtgaagcggctgtgcacatgccgagcgcaggaagactttgt 461
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1005 CGACCAGCTCTCGCGCGGGTCCGGTGTCTGCAAGGTGGCGCGCCGCAAGAACGACGT 946
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 462 ctctgagcctacctctgacctgtgcaagttgcattcatcaatgtttgtctctgatga 521
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 945 GCACATGGAAGACGTGCACCGCGCTGGGGCATCATGGGCATTCCTGGGCGAGCTTGACCG 886
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 522 gctaaagaacatgaagtcagcgtcagcaagaatgacctcgcctacaaagagggcagcaca 581
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 885 CGCGGGCTGCTGAAGACCGACGTGAGCAGCGTACACGCCGCCCGCAGCATGAGCGCCCGCT 826
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 582 gttctcggaagatggcagatccccagtcctccaggagt 622
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 825 GAACGAGTGGGACGTGGTCCGCAATGAGAACCACACGAGT 785
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 15
MXALON 4038 bp DNA BCT 03-FEB-1999
LOCUS
DEFINITION
Myxococcus xanthus gene for ATP-dependent protease La, complete
cvs.
ACCESSION D12923
VERSION D12923.1 GI:303711
KEYWORDS ATP-dependent protease La; lon.
SOURCE Myxococcus xanthus (strain:DZf1) DNA, clone:PMXL002.
ORGANISM Myxococcus xanthus
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
Komano,T.
Direct Submission
Submitted (17-AUG-1992) to the DDBJ/EMBL/GenBank databases. Teruya
Komano, Tokyo Metropolitan University, Biology; Minami-Ohsawa,
Hachioji, Tokyo 192-03, Japan
(E-mail:komano-teruya@cc.metro-u.ac.jp, Tel:0426-77-2568,
Fax:0426-77-2559)
REFERENCE 2 (bases 1 to 4038)
AUTHORS Tojo,N., Inouye,S. and Komano,T.
TITLE Cloning and nucleotide sequence of the Myxococcus xanthus lon gene:
indispensability of lon for vegetative growth
JOURNAL J. Bacteriol. 175 (8), 2271-2277 (1993)
MEDLINE 93224448
FEATURES
Location/Qualifiers
1. .4038
/organism="Myxococcus xanthus"
/strain="DZf1"
/db_xref="taxon:34"
/clone="PMXL002"
782. .787
-35_signal 804. .809
-10_signal
```

Wed May 24 09:19:39 2000

```
RBS      1056. .1061
gene      1069. .3522
          /gene="lon"
CDS       1069. .3522
          /gene="lon"
          /codon_start=1
          /transl_table=11
          /product="ATP-dependent protease La"
          /protein_id="BAA02307.1"
          /db_xref="GI:303712"
          /translation="MFEGRDCKKKAQKRLTVPLRLPRLDIIVPHMVVPLFVGRKSI
AALKDAMAHKGPDDKAVILLAAQKAKTNDPTDDIFHFGTLGHVLIQLPLPDGIVKV
LVEGVRAKVKKHPNDAPFMEVEVEVEQTEKTEVELEALVRSVSHSVFAFVKNKRI
PPMLQVASIDDPADRLADTVIAHLSLKNDKOALLETESPAKRLKLYELMQGEIEI
LQVEKKLRIRVKQMEKTOKEYYNQMOAIOKELGERDEFNEIOEIEEKLKNRMS
KEATLKVKELKLRMSPSAEATVVVRYNIDWIIISLPWDETQDRLDVTEAETVLE
DHYGLKPKLEILAVQOVLKLVKGLVLCFVGPVGKTSIARATGRKKEVRL
SLGGVRDEAETIRHRTYIGAMPFKLQSLKAGSNPNVFLLEIDKMSDTRGDSA
ALLEVDPQNHNDHYLDLDYLSKMFICTANTMHNIPGLQDMVEIRIAGTIE
PEKLSIARRYLIPIKEQANGLSOLKVDISDPALRTIIRHYTRESGVRSLEIRIGVFR
KIARDVLKNGKRDIDVRKAMKFLGTPRYRYGMAEADQVIGTGLAWTELGGEILT
TEATIMPGKGLIITGLIGVMOESAOAMSVYRSRAEREGIDRKVFENYDIHVHLP
GAIPKDGPSAGVTICTALVLSALRYLIRDVAMTGETLGRVLPISGLLKEKTLAAHR
AGIKTVLIPKANKDKLDIPLKIRKQIRIVFEFVDVLRALVLEKPEEFGRKPTTD
GGKLGTTTLPSPAPAPA"
```

BASE COUNT	765 a	1258 c	1410 g	605 t
ORIGIN				
Query Match	5.0%	Score 44.4;	DB 1;	Length 4038;
Best Local Similarity	44.7%	Pred. No. 0.26;		
Matches	217;	Conservative	Mismatches 266;	Indels 3; Gaps 1;
Qy	284	ccaggtgaatcaacagcagcagccacccagtagagatctatgagaagacagtagagg	343	
Db	1397	CCAGGTTGAAGAGTCCACCCCAACGACGCCTCTTCATGTCGAGGTGGAGAGGTGG	1456	
Qy	344	tgctggagcggaggtcaccagctcatgaagttcatgtattttcagcgaagccatcg	403	
Db	1457	AGGAGCAGCGGAGAGAGAGCGGTGGAGCGCTCGTTCGACGCTGATGCGAGTGGCCA	1516	
Qy	404	agcggttcgcagcaggtgaagcggctgtgccatgccagcgcaggaaggactttgtct	463	
Db	1517	TCGAGGCGCTTCGTCAAGCTCAACAGCGCATCCCGCCTGAGATGCTGATGCGAGTGGCCA	1576	
Qy	464	ctgaggcctacctctgaccttggaagttcatcaacatgtttgtctcctggatgac	523	
Db	1577	GCATCGAGACCCCGCGCGTGGCCGACACCATCGTCGCGCACCTCTCCCTG---AAGC	1633	
Qy	524	taagaacatgaagtgcagcgtcaagaatgaccactccgctacaagagggcagcacagt	583	
Db	1634	TGAACGACAACGAGCCCTGCTCGAGAGGAGTCCCGCGCCCAAGCGCCTGGAGAACTGT	1693	
Qy	584	tcctcggaagatggcagatcccccagttctatccaggagtcgcagaacctttccatgtcc	643	
Db	1694	ACGACCTGATCGAGGTCGAGATCGAGATTCTCCAGTGGAGAGAGATCCGCACCGCGG	1753	
Qy	644	tggccaacacacaggaataccacagtgctccacacgaacttgaagtgtatccaggct	703	
Db	1754	TCAAGAAGCAGATGGAGAGAGCCCAAGAGGAGTACTACCTGATGACGATGCAGGCCA	1813	
Qy	704	atgaggagctgctgctgacatgtcaacatctgtgtgattactacagagaacaagatgt	763	
Db	1814	TTTCAGAGAGAGTGGTGTGAGCGCGACGAGTTCAAGACGAGATTTCAGGAGATTGAGAGA	1873	
Qy	764	acctga	769	
Db	1874	AGCTGA	1879	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:59:52 ; Search time 282.38 Seconds
(without alignments)
791.209 Million cell updates/sec

Title: US-09-215-435-113

Perfect score: 893

Sequence: 1 gccatgatcagcagctcac.....aaaaaaaaaaaaaaaaaaaaa 893

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	187.8	21.0	287	X51880	Human secreted pro
2	40	4.5	459	X69348	Corn glycogenin cl
3	38.2	4.3	2502	1 Q80752	DNA-polymerase con
4	38.2	4.3	2502	1 Q65782	Thermus DNA polyme
5	38.2	4.3	2502	1 V63403	Consensus DNA sequ
6	38	4.3	963	1 T76626	Cleavase BN/thromb
7	38	4.3	963	1 V53941	Nucleotide sequenc
8	38	4.3	963	1 V63831	Cleavase BN/thromb
9	38	4.3	1600	1 T27686	Mutant Thermus aqu
10	38	4.3	1600	1 T70347	Synthesis deficien
11	38	4.3	1600	1 T76647	Taq gene 5' nuclea
12	38	4.3	1600	1 V53855	Nucleotide sequenc
13	38	4.3	1600	1 V65786	Thermus aquatiscus
14	38	4.3	1600	1 V63407	DNA sequence of a
15	38	4.3	1647	1 Q80742	5' Nuclease from T
16	38	4.3	1647	1 T27682	Mutant Thermus aqu
17	38	4.3	1647	1 T70344	Synthesis deficien
18	38	4.3	1647	1 T76644	Clone 4C encoding
19	38	4.3	1647	1 V65783	Thermus aquatiscus
20	38	4.3	1647	1 V63404	DNA sequence of a
21	38	4.3	2088	1 Q80743	5' Nuclease from T
22	38	4.3	2088	1 T27683	Mutant Thermus aqu
23	38	4.3	2088	1 T70345	Synthesis deficien
24	38	4.3	2088	1 T76645	Taq gene 5' nuclea
25	38	4.3	2088	1 V53853	Nucleotide sequenc
26	38	4.3	2088	1 V63784	Thermus aquatiscus
27	38	4.3	2088	1 V63405	DNA sequence of a
28	38	4.3	2274	1 Q23994	Mutant thermotabl
29	38	4.3	2364	1 Q23995	Mutant thermotabl
30	38	4.3	2499	1 Q23993	Mutant thermotabl
31	38	4.3	2499	1 Q24223	Modified Taq polym
32	38	4.3	2499	1 Q24224	Modified Taq polym
33	38	4.3	2499	1 Q24225	Modified Taq polym
34	38	4.3	2502	1 Q80746	MutTaq 5'nuclease

ALIGNMENTS

RESULT 1

X51880
ID X51880 standard; DNA; 287 BP.
AC X51880;
DT 22-JUN-1999 (first entry)
DE Human secreted protein 5; EST SEQ ID NO: 94.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductiv hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN W09906552-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1236.
PR 01-AUG-1997; US-905223.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153782/13.
DR P-PSDB; Y13080.
PT New isolated brain-derived nucleic acids - used to develop products
PT which may have cytokine, immune, regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
PS Claim 1; Page 244; 577pp; English.
CC X5187 to X52019 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12987 to Y13219,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductiv hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 287 BP; 66 A; 101 C; 71 G; 49 T;

Query Match 21.0%; Score 187.8; DB 1; Length 287;
Best Local Similarity 99.0%; Pred. No. 2.7e-42;
Matches 189; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gccatgatcagcagctcacctggaagatgcctgtccaacgtggacctgttggaagag 60

Db 95 GCCATGACCAGCAGCGTCACCTTGAAGATGCCCTGTCCAACGTGGACCTGCTTGAAGAG 154

QY 61 ctctccctcccccagaccagcgcatgatcgatcgagcctccactctctccatcatgtaccag 120

Db 155 CTTCCCTCCCTCCCGACCCAGCAGCCATGATGAGCCTTCTCTCCATCATGTACCAG 214

Mutant Thermus aqu
MutTaq 5'nuclease
Taq DNA polymerase
Nucleotide sequenc
Thermus aquatiscus
Synthesis deficien
T. aquatiscus DNA-p
Cleavase DN nuclea
Cleavase DA nuclea
Cleavase DV nuclea
Nucleotide sequenc

35 38 4.3 2502 1 T27681
36 38 4.3 2502 1 T70343
37 38 4.3 2502 1 T76643
38 38 4.3 2502 1 V53864
39 38 4.3 2502 1 V65795
40 38 4.3 2502 1 V63416
41 38 4.3 2505 1 Q80749
42 38 4.3 2505 1 T76627
43 38 4.3 2505 1 T76628
44 38 4.3 2505 1 T76629
45 38 4.3 2505 1 V53945

PS Disclosure: Page 88-90; 159pp; English.
CC The sequences of Taq, Tfl and Tth thermostable DNA-polymerases
CC are given in R64272-74, respectively. The consensus sequence of
CC these enzymes is given in R64275, and the encoding nucleotide
CC sequence in Q80752.
SQ Sequence 2502 BP; 403 A; 857 C; 839 G; 364 T;

Query Match 4.3%; Score 38.2; DB 1; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.47;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 121 gctaaattgacacaaatttgagacaggaagtgcattgttcacggcgattgcaaggtac 180
DB 215 GCTAACTTTGACACAACTTTGAGGACAGGAATGATTTGTTCACGGGCAATGCAAGGTAC 274

QY 181 attgagcaggc 191
DB 275 ATTGACCAAGGC 285

RESULT 2
ID V69348 standard; cDNA; 459 BP.
AC V69348;
DT 09-FEB-1999 (first entry)
DE Corn glycogenin clone crln.pk0033.g10 cDNA.
KW Glycogenin; water stress protein; starch biosynthesis; corn; rice;
KW wheat; plant; granules; branch chain; endosperm; ss.
OS Zea mays.
PN W09850553-A1.
PD 12-NOV-1998.
PF 06-MAY-1997; US-852615.
PR (DUPO) DU PONT DE NEMOURS & CO E I.
PA Lightner JE;
PI WPI; 98-610384/51.
DR P-PSDB; W82561.
DT Plant glycogenin and water stress proteins - used to alter plant
PT glycogenin and water stress protein expression.
PS Claim 2; Page 34; 57pp; English.
CC This sequence encodes a novel corn glycogenin protein from clone
CC crln.pk0033.g10. This protein is used in a method which alters plant
CC glycogenin and water stress protein expression. Manipulation of
CC glycogenin expression can be used to alter starch biosynthesis and
CC effects the number of starch granules in the endosperm of corn.
CC Overexpression or reduction of expression of genes encoding glycogenin
CC in corn, rice and wheat could be used to alter branch chain distribution
CC of the starch produced by these plants.
SQ Sequence 459 BP; 111 A; 131 C; 135 G; 81 T;

Query Match 4.5%; Score 40; DB 1; Length 459;
Best Local Similarity 58.3%; Pred. No. 0.075;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 399 catcagcgttctgacgcagatgagcgtgtgctgacgagcgcaggaagactt 458
DB 332 CAACGAGTGAACACTACAGCAAGTTCAGGCTGTGCGAGCTCACCAGTACGACAGGTGAT 391

QY 459 tgtcttgagcctacctctcctgaccccttggaagtttcacaaatgttctgtcctgga 518
DB 392 CTTATAGACGCGACCTCCTCATCTCTGAGGAACGTGCGACTTCTGTGCGCATGCGGA 451

RESULT 3
ID Q80752 standard; DNA; 2502 BP.
AC Q80752;
DT 19-JUL-1995 (first entry)
DE DNA-polymerase consensus sequence.
KW DNA-polymerase; DNAP; Thermus aquaticus; Taq; Thermus flavus; Tfl;
KW Thermus thermophilus; Tth; consensus sequence; ds.
OS Thermus spp.
PN W09429482-A.
PD 22-DEC-1994.
PF 06-JUN-1994; U06253.
PR 04-JUN-1993; US-073384.
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Brow MAD, Dahlberg JE, Lyamichev VI;
DR WPI; 95-036504/05.
DR P-PSDB; R64275.
PT 5' Nuclease(s) derived from thermostable DNA polymerase(s) - have
PT cleavage activity with reduced synthetic ability, used for
PT detection of specific target sequences.

PS Disclosure: Page 88-90; 159pp; English.
CC The sequences of Taq, Tfl and Tth thermostable DNA-polymerases
CC are given in R64272-74, respectively. The consensus sequence of
CC these enzymes is given in R64275, and the encoding nucleotide
CC sequence in Q80752.
SQ Sequence 2502 BP; 403 A; 857 C; 839 G; 364 T;

Query Match 4.3%; Score 38.2; DB 1; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.47;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 430 ctgtgccatccgagcgcaggaaggactttgtctctgagcctacatcctgacccctggc 489
DB 442 CAGCTCCTTTCCGACCGCATCGCGCTCTCCACCGCGGGGTACCTCATCACCCCGGG 501

QY 490 aagttcatcaactgtttgtctgtctctgagatgagtaagaacatgaatgcagcgtcaag 549
DB 502 TGGCTTTGGGAGAAAGTACGGCCTGAGGCGCGAGCAGTGGGTGGACTACCGGGCCCTGGCG 561

QY 550 aatgaccactccgcctacaaagaggcagcagcttctcgcggaagatggcagatcccccag 609
DB 562 GGGACCCCTCCGACAACTCCCGGGGTCAAGGGCATCGGGGAGAGACCGCCNGAAG 621

QY 610 tctatccaggagtcgcagaaacctttccatgttctcgtgcccacacacagatgatacccccag 669
DB 622 CTCCTCAGGAGTGGGGAGCCTCGAAAGCTCTCAAGAACCTTGGACCGGGGTGAAGCCC 681

QY 670 tgtctcaccagcaacttgaag 691
DB 682 GCCNTCCGGGAGAGATCCAGG 703

RESULT 4
ID V65782 standard; DNA; 2502 BP.
AC V65782;
DT 02-FEB-1999 (first entry)
DE Thermus DNA polymerase consensus DNA.
KW Nucleic acid detection; multiple sequential invasive cleavage;
KW DNA polymerase; ss.
OS Thermus aquaticus.
OS Thermus flavus.
OS Thermus thermophilus.
PN W09842873-A1.
PD 01-OCT-1998.
PF 24-MAR-1997; US-823516.
PR (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Brow MAD, Hall JG, Kwiatkowski RW, Lyamichev VI,
PI Mast AL, Vavra SH;
DR WPI; 98-557036/47.
DR P-PSDB; W79963.
PT Detecting target nucleic acid by sequence-specific cleavage of
PT complex with two specific oligonucleotides - used to detect
PT cytomegalovirus DNA
PS Example 2; Page 271-272; 524pp; English.
CC This is a consensus of nucleotides sequences (see V65779-81)
CC coding for the thermostable DNA polymerases (DNAP, see W79960-62)
CC of Thermus aquaticus (Taq), Thermus flavus and thermus thermophilus.
CC The invention relates to means for the detection and characterisation
CC of nucleic acid sequences, and variations in nucleic acid sequences.
CC It also relates to methods for forming a nucleic acid cleavage
CC structure on a target sequence and cleaving this structure in a
CC site-specific manner, preferably using a thermostable structure-
CC specific nuclease such as a modified Taq DNAP that has reduced
CC synthetic activity (see V65783-86). Cleavage of the cleavage
CC structure by the nuclease indicates the presence of specific
CC nucleic acid sequences or specific variants. The invention further
CC relates to methods for the separation of nucleic acid molecules
CC based on charge, methods for the detection of non-target cleavage
CC products via the formation of a complete and activated protein
CC binding region, and methods for the detection of nucleic acid from

CC directed cleavage reaction of the invention is an ideal direct
CC detection method that combines the advantages of the direct
CC detection assays (e.g. easy quantification and minimal risk of
CC carry-over contamination) with the specificity of a dual or
CC tri-oligonucleotide hybridisation assay.
SQ Sequence 963 BP; 179 A; 330 C; 312 G; 142 T;

Query Match 4.3%; Score 38; DB 1; Length 963;
Best Local Similarity 45.4%; Pred. No. 0.36;
Matches 137; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 430 ctgtccatgcagcagcaggaagactttgtctctgagcctacctcctgacctgtgc 489
DB 493 CAGTCCTTTCCGACCGCATCCAGCTCTCCACCCCGAGGGGTACCTCATCACC
QY 490 aagttcatcaacatgtttgtctctgagcctaaagaacatgaagtgcagcgtcaag 549
DB 553 TGGCTTTGGGAAAGTACGGCTGAGGCCCGACAGTGGGCCGACTACCGGGCCCTGACC 612
QY 550 aatgaccactccgcctacaaagaggcagcagttctctgcggaagatggcagatccccag 609
DB 613 GGGGACGAGTCCGACAACTTCCCGGGGTCAAGGGCATCGGGGAGAAACGCGGAGGAAG 672
QY 610 tctatccaggagtcagaaacctttccatgttctctgcccacacacagatcacccag 669
DB 673 CTTCTGGAGGAGTGGGGAGCTTGAAGCCCTCTCAGAACCTTGGACCGGCTGAAGCCC 732
QY 670 tgtctccaccagcaacttgaagtgtatccaggctatgagggagctgctggtgacattgtc 729
DB 733 GCATCCGGGAGAGATCTGSCCCACATGGAGCATCTGAAGCTCTCTCTGGGACCTGGCC 792
QY 730 aa 731
DB 793 AA 794

RESULT 7
V53941
ID V53941 standard; DNA; 963 BP.
AC V53941;
DT 21-DEC-1998 (first entry)
DE Nucleotide sequence of the cleavase BN/thrombin nuclease.
KW Cleavase; Taq gene; thermostable; structure-specific nuclease;
KW mutant; DNA polymerase; bacteria; fungi; protozoa; RNA virus;
KW hepatitis C virus; HCV; ds.
OS Synthetic.
FH Key
FT Location/Qualifiers
FT 1..963
FT CDS
FT /*tag= a
FT /product= "cleavase BN/thrombin nuclease"
PN WO9823774-A1.
PD 04-JUN-1998.
PE 26-NOV-1997; U21783.
PR 02-DEC-1996; US-758314.
PR 29-NOV-1996; US-757653.
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Kaiser MW, Lyamichev VI, Lyamicheva N;
DR WPI; 98-322748/28.
DR P-PSDB; W59937.
PT Thermostable structure-specific nuclease(s) derived from mutant DNA
PT polymerase(s) - useful for detecting mutant allele(s) or strains of
PT microorganisms
PS Claim 2; Pages 266-267; 472pp; English.
CC This is the nucleotide sequence encoding the cleavase BN/thrombin
CC nuclease protein, used in the method of the invention. In this
CC process thermostable structure-specific nucleases are derived from
CC mutant DNA polymerases, which can be used for detecting mutant
CC alleles or strains of microorganisms. The structure-specific
CC nucleases can be used in mixtures, compositions and kits to treat
CC nucleic acid, e.g. for detection of wild type and mutant alleles of
CC genes, for detection and/or identification of strains of
CC microorganisms such as bacteria, fungi, protozoa, especially for

CC detection of RNA viruses such as the hepatitis C virus (HCV).
SQ Sequence 963 BP; 179 A; 330 C; 312 G; 142 T;

Query Match 4.3%; Score 38; DB 1; Length 963;
Best Local Similarity 45.4%; Pred. No. 0.36;
Matches 137; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 430 ctgtgccatgcagcaggaagactttgtctctgagcctacctcctgacctgtgc 489
DB 493 CAGTCCTTTCCGACCGCATCCAGCTCTCCACCCCGAGGGGTACCTCATCACC
QY 490 aagttcatcaacatgtttgtctctgagcctaaagaacatgaagtgcagcgtcaag 549
DB 553 TGGCTTTGGGAAAGTACGGCTGAGGCCCGACAGTGGGCCGACTACCGGGCCCTGACC 612
QY 550 aatgaccactccgcctacaaagaggcagcagttctctgcggaagatggcagatccccag 609
DB 613 GGGGACGAGTCCGACAACTTCCCGGGGTCAAGGGCATCGGGGAGAAACGCGGAGGAAG 672
QY 610 tctatccaggagtcagaaacctttccatgttctctgcccacacacagatcacccag 669
DB 673 CTTCTGGAGGAGTGGGGAGCTTGAAGCCCTCTCAGAACCTTGGACCGGCTGAAGCCC 732
QY 670 tgtctccaccagcaacttgaagtgtatccaggctatgagggagctgctggtgacattgtc 729
DB 733 GCATCCGGGAGAGATCTGSCCCACATGGAGCATCTGAAGCTCTCTCTGGGACCTGGCC 792
QY 730 aa 731
DB 793 AA 794

RESULT 8
V65831
ID V65831 standard; DNA; 963 BP.
AC V65831;
DT 02-FEB-1999 (first entry)
DE Cleavase BN/thrombin nuclease DNA.
KW Nucleic acid detection; multiple sequential invasive cleavage;
KW DNA polymerase; nuclease; Cleavase BN; thrombin; ds.
OS Thermus aquaticus.
OS Synthetic.
PN WO9842873-A1.
PD 01-OCT-1998.
PE 24-MAR-1998; U05809.
PR 24-MAR-1997; US-823516.
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Brow MAD, Hall JG, Kwiatkowski RW, Lyamichev VI,
PI Mast AL, Vavra SH;
DR WPI; 98-557036/47.
DR P-PSDB; W79965.
PT Detecting target nucleic acid by sequence-specific cleavage of
PT complex with two specific oligonucleotides - used to detect
PT cytomagalovirus DNA
PS Example 27; Page 291-293; 524pp; English.
CC This nucleotide sequence codes for a thermostable nuclease (see
CC W79965), designated Cleavase BN/thrombin, that is derived from
CC Cleavase BN (see W65799) by having a thrombin recognition sequence
CC engineered into the region of Cleavase BN nuclease which is thought
CC to form the helical arch of the protein through which single-
CC stranded DNA that is cleaved is presumed to pass. Cleavase BN is
CC itself derived from the DNA polymerase (DNAP) of Thermus aquaticus
CC and has reduced synthetic activity compared to wild-type enzyme. The
CC Cleavase BN/thrombin has an enhanced ability to cleave circular
CC DNA and thus a reduced requirement for the presence of a free 5'
CC end. The invention relates to means for the detection and
CC characterisation of nucleic acid sequences, and variations in
CC nucleic acid sequences. It also relates to methods for forming a
CC nucleic acid cleavage structure on a target sequence and cleaving
CC this structure in a site-specific manner, preferably using a
CC thermostable structure-specific nuclease such as a modified Taq
CC DNAP that has reduced synthetic activity. Cleavage indicates the

CC activity. The invention relates to means for the detection and
CC characterisation of nucleic acid (NA) sequences and variations in
CC NA. It relates to methods for forming a NA cleavage structure on
CC a target sequence and cleaving the NA cleavage structure in a
CC site-specific manner. The 5' nuclease activity of various enzymes
CC (see W24210-13) is used to cleave the target-dependent cleavage
CC structure, thereby indicating the presence of specific NA sequences
CC or specific variations of them. 519 C; 553 G; 242 T;
SQ Sequence 1600 BP; 286 A; 519 C; 553 G; 242 T;

Query Match 4.3%; Score 38; DB 1; Length 1600;
Best Local Similarity 45.4%; Pred. No. 0.44;
Matches 137; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 430 ctgtgcatgcagcgaggaagacttctctgagcctacctctgacctctggc 489
DB 443 CAGCTCTTTTCGACCGCATCCAGTCTCCACCCCGAGGGGTACCTCATCCCGGCC 502
QY 490 aagttcatcaacatgttctgtctgagtgatgagctaaagaacatgaagtcagtc 549
DB 503 TGCGTTTGGGAAAGTACGCGCTGAGGCGCCGACAGTGGCGGCTGACC 562
QY 550 aatgacactccctacaagagcgacacagttcttcggaagatggcagatcccccag 609
DB 563 GGGGACGAGTCCGACAACTTCCCGGGTCAAGGGCATCGGGGAGAGGAG 622
QY 610 tctatccagagtcgcagaaaccttttccatgttctctgcccacacacagatcccccag 669
DB 623 CTTCCTGAGGAGTGGGGAGCCTTGGGAGCCCTCTCAAGAACCTGGACCGCTGAAGCCC 682
QY 670 tgtctccacagcaacttgaagtgtatcccgagctatgagagctgctgagctatgtc 729
DB 683 GCCATCCGGGAGAGATCCTGGCCCAATGAGGAGTCTCAAGCTCTCCTGGGACCTGGCC 742
QY 730 aa 731
DB 743 AA 744

RESULT 11
T76647
ID T76647 standard; DNA; 1600 BP.
AC T76647;
DT 14-APR-1998 (first entry)
DE Taq gene 5' nuclease clone 3F (Cleavase BB).
KW Nucleic acid cleavage; DNA cleavage; RNA cleavage; 5' nuclease;
KW Taq; DNA polymerase; Cleavase BB; ds.
OS Thermus aquaticus YT-1.
OS Synthetic.
PN WO9727214-A1.
PD 31-JUL-1997.
PF 22-JAN-1997; U01072.
PR 02-DEC-1996; US-759038.
PR 24-JAN-1996; US-599491.
PR 12-JUL-1996; US-682853.
PR 29-NOV-1996; US-756386.
PR 02-DEC-1996; US-758314.
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Brow MAD, Dahlberg JE, Hall JG, Kaiser MW, Lyamichev VI,
PI Olive DM, Prudent JR;
DR WPI; 97-393613/36.
PT Thermostable structure-specific nuclease(s) - used for detection and
PT characterisation of nucleic acid sequences and variations in nucleic
PT acid sequences
PS Example 2; Page 245; 457pp; English.
CC This DNA sequence, denoted clone 3F or Cleavase BB, comprises an
CC altered Thermus aquaticus DNA polymerase (Taq) gene in which
CC nucleotides 875-1778 of the wild-type gene coding sequence are
CC deleted. Mutant gene mutTaq (see T76643) was used as the starting
CC material for the construct. Cleavase BB is a thermostable
CC structure-specific nuclease preferred for use in nucleic acid
CC cleavage methods of the invention. Mutant genes (T76644-47)
CC were constructed in order to determine which portions of the Taq
CC polymerase domain can be altered without eliminating 5' nuclease

CC activity. The invention relates to means for the detection and
CC characterisation of nucleic acid (NA) sequences and variations in
CC NA. It relates to methods for forming a NA cleavage structure on
CC a target sequence and cleaving the NA cleavage structure in a
CC site-specific manner. The 5' nuclease activity of various enzymes
CC (see W24210-13) is used to cleave the target-dependent cleavage
CC structure, thereby indicating the presence of specific NA sequences
CC or specific variations of them. 519 C; 553 G; 242 T;
SQ Sequence 1600 BP; 286 A; 519 C; 553 G; 242 T;

Query Match 4.3%; Score 38; DB 1; Length 1600;
Best Local Similarity 45.4%; Pred. No. 0.44;
Matches 137; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 430 ctgtgcatgcagcgaggaagacttctctgagcctacctctgacctctggc 489
DB 443 CAGCTCTTTTCGACCGCATCCAGTCTCCACCCCGAGGGGTACCTCATCCCGGCC 502
QY 490 aagttcatcaacatgttctgtctcctgagtgatgagctaaagaacatgaagtcagtc 549
DB 503 TGCGTTTGGGAAAGTACGCGCTGAGGCGCCGACAGTGGCGGCTGACC 562
QY 550 aatgacactccctacaagagcgacacagttcttcggaagatggcagatcccccag 609
DB 563 GGGGACGAGTCCGACAACTTCCCGGGTCAAGGGCATCGGGGAGAGGAG 622
QY 610 tctatccagagtcgcagaaaccttttccatgttctctgcccacacacagatcccccag 669
DB 623 CTTCCTGAGGAGTGGGGAGCCTTGGGAGCCCTCTCAAGAACCTGGACCGCTGAAGCCC 682
QY 670 tgtctccacagcaacttgaagtgtatcccgagctatgagagctgctgagctatgtc 729
DB 683 GCCATCCGGGAGAGATCCTGGCCCAATGAGGAGTCTCAAGCTCTCCTGGGACCTGGCC 742
QY 730 aa 731
DB 743 AA 744

RESULT 12
V53855
ID V53855 standard; DNA; 1600 BP.
AC V53855;
DT 21-DEC-1998 (first entry)
DE Nucleotide sequence of clone 4F of the Taq gene mutant.
KW Clone 4D; Taq mutant gene; thermostable; structure-specific nuclease;
KW mutant DNA polymerase; bacteria; fungi; protozoa; RNA virus;
KW hepatitis C virus; HCV; ds.
OS Thermus sp.
PN WO9823774-A1.
PD 04-JUN-1998.
PF 26-NOV-1997; U21783.
PR 02-DEC-1996; US-758314.
PR 29-NOV-1996; US-757653.
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Kaiser MW, Lyamichev VI, Lyamicheva N;
PI WPI; 98-332748/28.
DR Thermostable structure-specific nuclease(s) derived from mutant DNA
PT polymerase(s) - useful for detecting mutant allele(s) or strains of
PT microorganisms
PS Example 2; Page 238; 472pp; English.
CC This is the nucleotide sequence of a clone of a mutant Taq gene,
CC used in the method of the invention. In this process thermostable
CC structure-specific nucleases are derived from mutant DNA polymerases,
CC which can be used for detecting mutant alleles or strains of
CC microorganisms. The structure-specific nucleases can be used in
CC mixtures, compositions and kits to treat nucleic acid, e.g. for
CC detection of wild type and mutant alleles of genes, for detection
CC and/or identification of strains of microorganisms such as bacteria,
CC fungi, protozoa, especially for detection of RNA viruses such as the
CC hepatitis C virus (HCV).
SQ Sequence 1600 BP; 286 A; 519 C; 553 G; 242 T;


```
Query Match 4.3%; Score 38; DB 1; Length 1600;
Best Local Similarity 45.4%; Pred. No. 0.44;
Matches 137; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 430 ctgtgcattgcccagcagcaggaagactttgtctctgagcctacctctgaccccttggc 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 CAGCTCTTTCCGACGCATCCACGTCCTCCACCCGAGGGGTACCTCATCACCCCGGCC 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 490 aagtcatcaacatgttctgtctgagtgatgagcgttaaaagaacatgaagtcagcgtcaag 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 TGCTTTGGGAAAGTACGGCCTGAGCCCGACCTGAGGGCCGACTACCGGGCCCTGACC 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 550 aatgacactccgcctacaaagagcagcacagtcttcctgcggaagatggcagatcccccag 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 GGGGACGAGTCCGACAACTTCCCGGGTCAAGGGCATCGGGGAGAAGACGGCGAGGAAG 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 610 tctatccaggagtcgcagaacctttccatgttctctgcccacacacagatcaccacag 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 CTCTGGAGGAGTGGGGAGCCTTGGAGCCCTCTCAAGAACCTGGACCGGCTGAAGCCC 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 670 tgtctccacagcaacttgaagtgtatccaggtatgagagctgtgctgacattgtc 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 683 GCCATCGGGAGAAGATCCTTGGCCACATGGACGATCTGAAAGCTCTCTCTGGGACCTGGCC 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 730 aa 731
    ||
Db 743 AA 744

RESULT 13
V65786
ID V65786 standard; DNA; 1600 BP.
AC V65786;
DE Thermus aquaticus nuclease clone 3F (Cleavase BB) DNA.
KW Nucleic acid detection; multiple sequential invasive cleavage;
KW DNA polymerase; nuclease; Cleavase BB; ds.
OS Thermus aquaticus strain YT-1.
OS Synthetic.
PN W09842873-AL.
PD 01-OCT-1998.
PF 24-MAR-1997; US-823516.
PR 24-MAR-1997; US-823516.
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Brow MAD, Hall JG, Kwiatkowski RW, Lyamichev VI,
PI Mast AL, Vavra SH;
PR WPI; 98-557036/47.
PT Detecting target nucleic acid by sequence-specific cleavage of
PT complex with two specific oligonucleotides - used to detect
PT cytomegalovirus DNA
PS Example 2; Page 278; 524pp; English.
CC This nucleotide sequence (clone 3F) codes for a thermostable
CC nuclease (Cleavase BB) derived from the DNA polymerase (DNAP) of
CC Thermus aquaticus (Taq). In comparison to the wild-type Taq DNAP
CC sequence (see V65779), it contains an in-frame deletion of 903
CC nucleotides. The invention relates to means for the detection and
CC characterisation of nucleic acid sequences, and variations in
CC nucleic acid sequences. It also relates to methods for forming a
CC nucleic acid cleavage structure on a target sequence and cleaving
CC this structure in a site-specific manner, preferably using a
CC thermostable structure-specific nuclease such as a modified Taq
CC DNAP that has reduced synthetic activity (see V65783-86). Cleavage
CC of the cleavage structure by the nuclease indicates the presence of
CC specific nucleic acid sequences or specific variants. The invention
CC further relates to methods for the separation of nucleic acid
CC molecules based on charge, methods for the detection of non-target
CC cleavage products via the formation of a complete and activated
CC protein binding region, and methods for the detection of nucleic
CC acid from various viruses (e.g. human cytomegalovirus) in a sample.
CC The method amplifies the detection molecule rather than the target
CC itself, is less subject to contamination than exponential
```

```
CC amplification processes, and allows many targets to be analysed in
CC a single reaction.
SQ Sequence 1600 BP; 286 A; 519 C; 553 G; 242 T;

Query Match 4.3%; Score 38; DB 1; Length 1600;
Best Local Similarity 45.4%; Pred. No. 0.44;
Matches 137; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 430 ctgtgcattgcccagcagcaggaagactttgtctctgagcctacctctgaccccttggc 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 CAGCTCTTTCCGACGCATCCACGTCCTCCACCCGAGGGGTACCTCATCACCCCGGCC 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 490 aagtcatcaacatgttctgtctgagtgatgagcgttaaaagaacatgaagtcagcgtcaag 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 TGCTTTGGGAAAGTACGGCCTGAGCCCGACCTGAGGGCCGACTACCGGGCCCTGACC 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 550 aatgacactccgcctacaaagagcagcacagtcttcctgcggaagatggcagatcccccag 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 GGGGACGAGTCCGACAACTTCCCGGGTCAAGGGCATCGGGGAGAAGACGGCGAGGAAG 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 610 tctatccaggagtcgcagaacctttccatgttctctgcccacacacagatcaccacag 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 CTCTGGAGGAGTGGGGAGCCTTGGAGCCCTCTCAAGAACCTGGACCGGCTGAAGCCC 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 670 tgtctccacagcaacttgaagtgtatccaggtatgagagctgtgctgacattgtc 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 683 GCCATCGGGAGAAGATCTTGGCCACATGGACGATCTGAAAGCTCTCTCTGGGACCTGGCC 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 730 aa 731
    ||
Db 743 AA 744

RESULT 14
V63407
ID V63407 standard; DNA; 1600 BP.
AC V63407;
DT 26-JAN-1999 (first entry)
DE DNA sequence of a thermostable 5' nuclease derived from Taq polymerase.
KW Thermostable DNA polymerase; nucleic acid detection; Cleavase BB;
KW thermostable 5' nuclease; ds.
OS Synthetic.
OS Thermus aquaticus.
PN US5837450-A.
PD 17-NOV-1998.
PF 06-JUN-1995; 471066.
PR 06-JUN-1994; US-254359.
PR 07-DEC-1992; US-986330.
PR 04-JUN-1993; US-073384.
PR 06-JUN-1995; US-471066.
PA (THIR-) THIRD WAVE TECHNOLOGIES INC.
PI Brow MAD, Dahlberg JE, Lyamichev VI;
PI WPI; 99-023438/02.
PT Detection of target nucleic acid molecules - uses modified
PT thermostable enzymes with specific cleavage activity to create
PT specific detection products from oligonucleotide(s) and target
PT hybridisation
PS Claim 10; Columns 79-82; 91pp; English.
CC The present sequence represents a thermostable 5' nuclease derived from
CC a thermostable polymerase modified to have reduced synthetic activity,
CC where the 5' nuclease is capable of cleaving a linear nucleic acid
CC duplex structure to create a single, single-stranded cleavage product.
CC The nuclease, designated Cleavase BB, is used in a method for detecting
CC the presence of a nucleic acid molecule. The method is used for the
CC specific detection of nucleic acid sequences, via a cleavage-based
CC procedure, but without the need for amplification of target sequences.
CC Thermostable polymerases, altered to have nuclease, but not polymerase
CC activity are preferably used due to their specificity. The cleavage
CC product specifically formed is detected, preferably by the use of
CC radioactively labelled oligonucleotides. These can be used in
CC e.g. forensic testing or paternity determination.
SQ Sequence 1600 BP; 286 A; 519 C; 553 G; 242 T;
```

Query Match

Best Local Similarity

Matches 137; Conservative

4.3%;

45.4%;

0;

Score 38;

Pred. No. 0.44;

Mismatches 165;

DB 1;

Length 1600;

Indels

0;

Gaps

0;

QY

430

ctgtgcatccgagcgaggaagacttctctgagcctaccctcctgacccctggc

489

DB

443

CAGCTCCTTCGACCGGATCCACCTCTCCACCCGAGGGGTACCTCATCACCCTGGGCC

502

QY

490

aagttcatcaacatgttctgtctgctgagatgagctaaagacatgaagtcagcgtcaag

549

DB

503

TGGCTTTGGGAAAGTACGGCTGAGGCCCGGACCACTGGCGCGACTACCGGGCCCTGACC

562

QY

550

aatgaccactccgcataaagggcagacagttcctctgagcctaccctcctgacccctggc

609

DB

563

GGGACGAGTCCGACACCTTCCCGGGGTGAAGGCGATCGGGGAGAACGGCGAGGAAG

622

QY

610

tctatccagagtcgcagaacacttccatgttccctggcgaacacacagagatcacccag

669

DB

623

CTTCTGGAGAGTGGGGAGCTTGAAGCCCTCTCAGAACCTTGGACCGGCTGAAGGCC

682

QY

670

tgtctccaccagcaacttgaagtgatcccaaggctatgaggagctgctggctgacattgtc

729

DB

683

GCCATCCGGGAGAGATCTGGCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCC

742

QY

730

aa 731

DB

743

AA 744

Search completed: May 24, 2000, 00:00:04

Job time: 10690 sec

RESULT 15

Q80742

AC Q80742 standard; DNA; 1647 BP.

ID Q80742;

DE 19-JUL-1995 (first entry)

DT 5' Nuclease from Taq DNAP.

KE DNA-polymerase; DNAP; Taq; DNA cleavage; RNA cleavage; 5' nuclease; ds.

OS Thermus aquaticus YT-1.

PN W09429482-A.

PD 22-DEC-1994.

PF 06-JUN-1994; U06253.

PR 04-JUN-1993; US-073384.

PA (THIR-) THIRD WAVE TECHNOLOGIES INC.

PI Brow MAD, Dahlberg JE, Lyamichev VI;

DR WPI; 95-036504/05.

PT 5' Nuclease(s) derived from thermostable DNA polymerase(s) - have

PT cleavage activity with reduced synthetic ability, used for

PT detection of specific target sequences.

PS Claim 5; Page 92-93; 159pp; English.

CC The Taq DNA-polymerase (DNAP) gene was amplified by PCR. Amplified

CC fragments were ligated into pTTQ18 vector, which contains the

CC hybrid trp-lac (tac) promoter. An amplification/selection

CC protocol was used to isolate clone 4B containing a mutated Taq DNAP

CC gene (mutTaq) (sequence given in Q80746) having normal 5' nuclease

CC activity but less than 1% of the wt Taq DNAP activity. mutTaq was

CC digested with PstI and religated to create an 894-nucleotide deletion.

CC The DNA sequence of the resulting 5' nuclease is given in Q80742.

CC Sequence 1647 BP; 273 A; 563 C; 559 G; 252 T;

QY

Query Match

Best Local Similarity

Matches 137; Conservative

4.3%;

45.4%;

0;

Score 38;

Pred. No. 0.45;

Mismatches 165;

DB 1;

Length 1647;

Indels

0;

Gaps

0;

QY

430

ctgtgcatccgagcgaggaagacttctctgagcctaccctcctgacccctggc

489

DB

442

CAGCTCCTTTCCGACCGGATCCACCTCTCCACCCGAGGGGTACCTCATCACCCTGGGCC

501

QY

490

aagttcatcaacatgttctgtctgctgagatgagctaaagacatgaagtcagcgtcaag

549

DB

502

TGGCTTTGGGAAAGTACGGCTGAGGCCCGGACCACTAGTGGGCCCGACTACCGGGCCCTGACC

561

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:54:38 ; Search time 203.6 seconds
(without alignments)
570.121 Million cell updates/sec

Title: US-09-215-435-113
Perfect score: 893
Sequence: 1 gccatgacgcgcgtcac.....aaaaaaaaaaaaaaaa 893

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2.6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2.6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2.6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.2	4.3	2502	1 US-08-073-384C-7	Sequence 7, Appli
2	38.2	4.3	2502	1 US-08-254-359A-7	Sequence 7, Appli
3	38.2	4.3	2502	2 US-08-483-043-7	Sequence 7, Appli
4	38.2	4.3	2502	2 US-08-481-238-7	Sequence 7, Appli
5	38.2	4.3	2502	3 US-08-471-066B-7	Sequence 7, Appli
6	38.2	4.3	2502	3 US-08-484-956-7	Sequence 7, Appli
7	38.2	4.3	2502	3 US-08-757-653-7	Sequence 7, Appli
8	38.2	4.3	2502	3 US-08-599-491-7	Sequence 7, Appli
9	38.2	4.3	2502	4 US-08-756-386-7	Sequence 7, Appli
10	38.2	4.3	2502	4 US-08-823-516-7	Sequence 7, Appli
11	38.2	4.3	2502	5 US-08-682-853A-7	Sequence 7, Appli
12	38	4.3	963	3 US-08-757-653-162	Sequence 162, App
13	38	4.3	963	3 US-08-823-516-60	Sequence 60, Appl
14	38	4.3	1600	1 US-08-073-384C-12	Sequence 12, Appl
15	38	4.3	1600	1 US-08-254-359A-12	Sequence 12, Appl
16	38	4.3	1600	2 US-08-483-043-12	Sequence 12, Appl
17	38	4.3	1600	2 US-08-481-238-12	Sequence 12, Appl
18	38	4.3	1600	3 US-08-471-066B-12	Sequence 12, Appl
19	38	4.3	1600	3 US-08-484-956-12	Sequence 12, Appl
20	38	4.3	1600	3 US-08-757-653-12	Sequence 12, Appl
21	38	4.3	1600	3 US-08-599-491-12	Sequence 12, Appl
22	38	4.3	1600	3 US-08-756-386-12	Sequence 12, Appl
23	38	4.3	1600	4 US-08-823-516-12	Sequence 12, Appl
24	38	4.3	1600	5 US-08-682-853A-12	Sequence 12, Appl
25	38	4.3	1647	1 US-08-073-384C-9	Sequence 9, Appli
26	38	4.3	1647	1 US-08-254-359A-9	Sequence 9, Appli
27	38	4.3	1647	2 US-08-483-043-9	Sequence 9, Appli

28	4.3	1647	2	US-08-481-238-9	Sequence 9, Appli
29	4.3	1647	3	US-08-471-066B-9	Sequence 9, Appli
30	4.3	1647	3	US-08-484-956-9	Sequence 9, Appli
31	4.3	1647	3	US-08-757-653-9	Sequence 9, Appli
32	4.3	1647	3	US-08-599-491-9	Sequence 9, Appli
33	4.3	1647	4	US-08-756-386-9	Sequence 9, Appli
34	4.3	1647	4	US-08-823-516-9	Sequence 9, Appli
35	4.3	1647	5	US-08-682-853A-9	Sequence 9, Appli
36	4.3	2088	1	US-08-073-384C-10	Sequence 10, Appl
37	4.3	2088	1	US-08-254-359A-10	Sequence 10, Appl
38	4.3	2088	2	US-08-483-043-10	Sequence 10, Appl
39	4.3	2088	2	US-08-481-238-10	Sequence 10, Appl
40	4.3	2088	3	US-08-471-066B-10	Sequence 10, Appl
41	4.3	2088	3	US-08-484-956-10	Sequence 10, Appl
42	4.3	2088	3	US-08-757-653-10	Sequence 10, Appl
43	4.3	2088	3	US-08-599-491-10	Sequence 10, Appl
44	4.3	2088	4	US-08-756-386-10	Sequence 10, Appl
45	4.3	2088	4	US-08-823-516-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-073-384C-7
; Sequence 7, Application US/08073384C
; Patent No. 5541311
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,384C
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-073-384C-7

Query Match 4.3%; Score 38.2; DB 1; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

		Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;			
Qy	430	ctgtgcaatgccgagcaggaagtgttctctagcctacctcctgacccctggc	489		
Db	442	CAGTCTCTTCCGACGATCGCGTCTCACCCTGAGGGGTACCTCATCACCCCGCG	501		
Qy	490	aagtcatcaacatgttctgtctgagctaaagaacatgaagtgcagctcaag	549		
Db	502	TGGCTTTGGGAGAGTACGGCTGAGCGGAGCAGTGGGTGGACTACCGGGCCCTGGCG	561		
Qy	550	aatgaccactccgctacaagagggcagcagttctctggaagatggcagatcccccag	609		
Db	562	GGGACCCCTCCGACACCTCCCGGGGTCAAGGGCATCGGGGAGAGACCGCCNGAAG	621		
Qy	610	tctatcaggagtgcagaactttccatgttctctgccaacacacacagatcacccag	669		
Db	622	CTCCTCNAGGAGTGGGGAGCTTGAAACCTCTCAAGAACCTTGAGCCGGTGAAGCCC	681		
Qy	670	tgctccaccagcaacttgaag	691		
Db	682	GCNTCCGGGAGAGATCCAGG	703		
RESULT 3					
US-08-483-043-7					
; Sequence 7, Application US/08483043					
; Patent No. 5691142					
; GENERAL INFORMATION:					
; APPLICANT: DAHLBERG, JAMES E.					
; APPLICANT: LYAMICHEV, VICTOR I.					
; TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE					
; TITLE OF INVENTION: DNA POLYMERASE					
; NUMBER OF SEQUENCES: 40					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL					
; STREET: 220 MONTGOMERY STREET, SUITE 2200					
; CITY: SAN FRANCISCO					
; STATE: CALIFORNIA					
; COUNTRY: UNITED STATES OF AMERICA					
; ZIP: 94104					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Floppy disk					
; COMPUTER: IBM PC compatible					
; OPERATING SYSTEM: PC-DOS/MS-DOS					
; SOFTWARE: Patent In Release #1.0, Version #1.25					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/08/254,359A					
; FILING DATE:					
; CLASSIFICATION: 435					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US 08/073,384					
; FILING DATE: 06-JUN-1993					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US 07/986,330					
; FILING DATE: 07-DEC-1992					
; ATTORNEY/AGENT INFORMATION:					
; NAME: CARROLL, PETER G.					
; REGISTRATION NUMBER: 32,837					
; REFERENCE/DOCKET NUMBER: FORS-01000					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: (415) 705-8410					
; TELEFAX: (415) 397-8338					
; INFORMATION FOR SEQ ID NO: 7:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 2502 base pairs					
; TYPE: nucleic acid					
; STRANDEDNESS: single					
; TOPOLOGY: linear					
; MOLECULE TYPE: DNA (genomic)					
US-08-254-359A-7					
Query Match 4.3%; Score 38.2; DB 1; Length 2502;					
Best Local Similarity 46.2%; Pred. No. 0.25;					

Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 430 ctgtgccatccgcagcaggaagacttctctgtgagcctacacctgacccttggc 489
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 442 CAGCTCTTTCACGCGCATCGCGCTCTCCACCCGAGGAGGTACTCATCACCCCGGCG 501
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 490 aagtcatcaacatgttctgtctgtggtgagctaaagaacatgaagtgcagcgtcaag 549
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 502 TGGCTTTGGGAGAAGTACGGCCTGAGCGCGGAGCAGTGGGTGGACTACCGGGCCCTGGCG 561
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 550 aatgacactccgcctacagaagggcagacacagttctctgcggaagatgcagatccccag 609
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 562 GGGGACCCCTCCGACAACTCCCGGGGTCAAGGGATCCGGGAGAGAACCGCCCGAAG 621
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 610 tctatcaggagtcgcagaaactttctctgtctgtggtgagcctaaagaacatgaagtgcagcgtcaag 669
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 622 CTCTCNAGAGTGGGGAGCCTGGAAACCTCTCAAGNACCTGACCGGGGTGAAGCCC 681
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 670 tgtctccaccagcaacttgaag 691
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 682 GCNCTCGGGAGAAGATCCAGG 703
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 4

```
US-08-481-238-7
; Sequence 7, Application US/08481238
; Patent No. 5795763
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,238
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-481-238-7
```

Query Match 4.3%; Score 38.2; DB 2; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 430 ctgtgccatccgcagcaggaagacttctctgtgagcctacacctgacccttggc 489
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 442 CAGCTCTTTCACGCGCATCGCGCTCTCCACCCCGAGGGTACTCATCACCCCGGCG 501
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 490 aagtcatcaacatgttctgtctgtggtgagcctaaagaacatgaagtgcagcgtcaag 549
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 502 TGGCTTTGGGAGAAGTACGGCCTGAGCGCGGAGCAGTGGGTGGACTACCGGGCCCTGGCG 561
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 550 aatgacactccgcctacagaagggcagacacagttctctgcggaagatgcagatccccag 609
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 562 GGGGACCCCTCCGACAACTCCCGGGGTCAAGGGATCCGGGAGAGAACCGCCCGAAG 621
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 610 tctatcaggagtcgcagaaactttctctgtctgtggtgagcctaaagaacatgaagtgcagcgtcaag 669
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 622 CTCTCNAGAGTGGGGAGCCTGGAAACCTCTCAAGNACCTGACCGGGGTGAAGCCC 681
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 670 tgtctccaccagcaacttgaag 691
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 682 GCNCTCGGGAGAAGATCCAGG 703
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 5

```
US-08-471-066B-7
; Sequence 7, Application US/08471066B
; Patent No. 5837450
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: 5' Nucleases Derived From Thermostable
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,066B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/254,359
; FILING DATE: 06-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,384
; FILING DATE: 04-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-01800
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-471-066B-7
```

Query Match 4.3%; Score 38.2; DB 3; Length 2502;

```
;
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-484-956-7

Query Match      4.3%; Score 38.2; DB 3; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Qy 430 ctgtgcatccagcagcaggaagacttctctgagcctacacctcctgaccttggc 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 CAGCTCTTTCCGACCGCATCGCCGTCCTCCACCCGAGGGGTACCTCATCACCCGGCG 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 490 aagttcatcaacatgttctgtctgagcagcagcagcagcagcagcagcagcag 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 TGGCTTTGGGAGAGTAGCGGCTGAGCGCGGAGCAGTGGGTGGACTACCGGGCCCTGGCG 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 550 aatgaccactccgctacaaagagggcagcagcagcagcagcagcagcagcagcag 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 GGGGACCCCTCCGACACACCTCCCGGGGTCAAGGGCATCGGGGAGAGACCGCCNGAAG 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 610 tctatccaggagtcgcagaacaccttctctgagcagcagcagcagcagcagcagcag 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 622 CTCTCNAGGAGTGGGGAGGCTTGGAAACCTCTCTCAAGAACCTCTGGAACCTGGACCGG 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 670 tgtctccaccagcaacttgaag 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 682 GCNCTCCGGGAGAGATCCAGG 703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-08-757-653-7
; Sequence 7, Application US/08757653
; Patent No. 5843669
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Lyamichev, Natasha
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,653
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02565
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; Gaps: 0;

;
; Best Local Similarity 46.2%; Pred. No. 0.25;
; Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Qy 430 ctgtgcatccagcagcaggaagacttctctgagcctacacctcctgaccttggc 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 CAGCTCTTTCCGACCGCATCGCCGTCCTCCACCCGAGGGGTACCTCATCACCCGGCG 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 490 aagttcatcaacatgttctgtctgagcagcagcagcagcagcagcagcagcag 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 TGGCTTTGGGAGAGTAGCGGCTGAGCGCGGAGCAGTGGGTGGACTACCGGGCCCTGGCG 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 550 aatgaccactccgctacaaagagggcagcagcagcagcagcagcagcagcagcag 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 GGGGACCCCTCCGACACCTCCCGGGGTCAAGGGCATCGGGGAGAGACCGCCNGAAG 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 610 tctatccaggagtcgcagaacaccttctctgagcagcagcagcagcagcagcagcag 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 622 CTCTCNAGGAGTGGGGAGGCTTGGAAACCTCTCTCAAGAACCTCTGGAACCTGGACCGG 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 670 tgtctccaccagcaacttgaag 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 682 GCNCTCCGGGAGAGATCCAGG 703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-08-484-956-7
; Sequence 7, Application US/08484956
; Patent No. 5843654
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; APPLICANT: OLDENBURG, MARY C.
; APPLICANT: HEISLER, LAURA
; TITLE OF INVENTION: DETECTION OF p53 MUTATIONS
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSHOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,956
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,601
; FILING DATE: 09-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,164
; FILING DATE: 09-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/254,359
; FILING DATE: 06-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,384
; FILING DATE: 04-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL J, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; 
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-653-7

Query Match 4.3%; Score 38.2; DB 3; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 430 ctgtgccatgcgagcgaggaagactttgtctcttgagggcctacctctctgaccttggc 489
DB 442 CAGCTCCTTTCCGACCGCATCCGCTCCACCCGAGGGGTACCTCATCACCCGGCG 501
QY 490 aagttcatacaatgtttgtctctctgagctaaagaacatgaagtgcagcgtcaag 549
DB 502 TGGCTTTGGGAGAGTACGGCTGAGCGCGGAGCAGTGGGTGACTACCGGCCCTGGCG 561
QY 550 aatgaccactccgcctacaaagggcgagcacagttcctcggaagatggcagatccccag 609
DB 562 GGGGACCCCTCCGACAACTCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCNGAAG 621
QY 610 ttatccagagtcgcagaaactttccatgttcttgcccaacacacagagatccccag 669
DB 622 CTCCTCAGGAGTGGGGAGCCTGGAACCTCTCAAGAACCTGGACCGGGTGAAGCCC 681
QY 670 tgtctccaccagcaacttgaag 691
DB 682 GCGNTCCGGGAGAGATCCAGG 703

RESULT 8

US-08-599-491-7
Sequence 7, Application US/08599491
Patent No. 5846717
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
APPLICANT: GROTELUESCHEN HALL, JEFF S.
APPLICANT: LYAMICHEV, VICTOR
APPLICANT: OLIVE, DAVID M.
APPLICANT: PRUDENT, JAMES R.
TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY
INVASIVE CLEAVAGE
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,491
FILING DATE: 23-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: P-40,027
REFERENCE/DOCKET NUMBER: FORS-01802
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-599-491-7

Query Match 4.3%; Score 38.2; DB 3; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 430 ctgtgccatgcgagcgaggaagactttgtctcttgagggcctacctctctgaccttggc 489
DB 442 CAGCTCCTTTCCGACCGCATCCGCTCCACCCGAGGGGTACCTCATCACCCGGCG 501
QY 490 aagttcatacaatgtttgtctctctgagctaaagaacatgaagtgcagcgtcaag 549
DB 502 TGGCTTTGGGAGAGTACGGCTGAGCGCGGAGCAGTGGGTGACTACCGGCCCTGGCG 561
QY 550 aatgaccactccgcctacaaagggcgagcacagttcctcggaagatggcagatccccag 609
DB 562 GGGGACCCCTCCGACAACTCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCNGAAG 621
QY 610 ttatccagagtcgcagaaactttccatgttcttgcccaacacacagagatccccag 669
DB 622 CTCCTCAGGAGTGGGGAGCCTGGAACCTCTCAAGAACCTGGACCGGGTGAAGCCC 681
QY 670 tgtctccaccagcaacttgaag 691
DB 682 GCGNTCCGGGAGAGATCCAGG 703

RESULT 9

US-08-756-386-7
Sequence 7, Application US/08756386
Patent No. 5985557
GENERAL INFORMATION:
APPLICANT: Prudent, James R.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamichev, Victor I.
TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/756,386
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02564
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid

us-09-215-435-113.rni

Wed May 24 09:19:40 2000

```
;
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-756-386-7

Query Match      4.3%; Score 38.2; DB 4; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 430 ctgtgcatgccagcagcaggaagactttgtctctgagcctacacctcctgaccttggc 489
Db 442 CAGCTCCTTTCCGACGATCGCGTCTCCACCCGAGGGTACCTCATCACCCCGCGG 501
QY 490 aagttcatcaaatgttctgtctctgagcctaaagaacatgaagtcagcgtcaag 549
Db 502 TGGCTTTGGGAGAGTACGGCTGAGCGCGGATCGGGGTCAAGGGCATCGGGGAGAGACCCCGCGG 561
QY 550 aatgaccactccgctacaaagaggcagcagcttctctgagcctcctgagcagatccccag 609
Db 562 GGGGACCCCTCGACAACTCCCGGGGTCAAGGGCATCGGGGAGAGACCCCGCGG 621
QY 610 tctatccaggagtcgagaacatttccatgttctctgagcctcctgagcagatccccag 669
Db 622 CTCCTCNAGGAGTGGGGAGCGCTCGAAACCTCTCAAGAACCTTGGACCGGGTGAAGCCC 681
QY 670 tgtctccaccagcaacttgaag 691
Db 682 GCNTCCGGGAGAGATCCAGG 703

RESULT 10
US-08-823-516-7
; Sequence 7, Application US/08823516
; Patent No. 5994069
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichiev, Victor I.
; APPLICANT: Mast, Andrea L.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
; TITLE OF INVENTION: Sequential Invasive Cleavages
; NUMBER OF SEQUENCES: 163
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,516
; FILING DATE: 24-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01072
; FILING DATE: 21-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/759,038
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/758,314
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/756,386
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853

;
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: F0RS-02736
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 703-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-823-516-7

Query Match      4.3%; Score 38.2; DB 4; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 430 ctgtgcatgccagcagcaggaagactttgtctctgagcctacacctcctgaccttggc 489
Db 442 CAGCTCCTTTCCGACGATCGCGTCTCCACCCGAGGGTACCTCATCACCCCGCGG 501
QY 490 aagttcatcaaatgttctgtctctgagcctaaagaacatgaagtcagcgtcaag 549
Db 502 TGGCTTTGGGAGAGTACGGCTGAGCGCGGATCGGGGTCAAGGGCATCGGGGAGAGACCCCGCGG 561
QY 550 aatgaccactccgctacaaagaggcagcagcttctctgagcctcctgagcagatccccag 609
Db 562 GGGGACCCCTCGACAACTCCCGGGGTCAAGGGCATCGGGGAGAGACCCCGCGG 621
QY 610 tctatccaggagtcgagaacatttccatgttctctgagcctcctgagcagatccccag 669
Db 622 CTCCTCNAGGAGTGGGGAGCGCTCGAAACCTCTCAAGAACCTTGGACCGGGTGAAGCCC 681
QY 670 tgtctccaccagcaacttgaag 691
Db 682 GCNTCCGGGAGAGATCCAGG 703

RESULT 11
US-08-823-516-7
; Sequence 7, Application US/08682853A
; Patent No. 6001567
; GENERAL INFORMATION:
; APPLICANT: Brow, Mary Ann D.
; APPLICANT: Grotelueschen Hall, Jeff S.
; APPLICANT: Lyamichiev, Victor
; APPLICANT: Olive, David M.
; APPLICANT: Prudent, James R.
; TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY
; TITLE OF INVENTION: INVADER-DIRECTED CLEAVAGE
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,853A
```


/ FILING DATE: 12-JUL-1996
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/599,491
/ FILING DATE: 24-JAN-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ingolia, Diane E.
/ REGISTRATION NUMBER: 40,027
/ REFERENCE/DOCKET NUMBER: FORS-02306
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 705-8410
/ TELEFAX: (415) 397-8338
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2502 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-682-853A-7

Query Match 4.3%; Score 38.2; DB 5; Length 2502;
Best Local Similarity 46.2%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 430 ctgtgcacatccgcagcaggaagactttgtctctgagccctacccctgacctgtgc 489
Db 442 CAGCTCTTTCCGACCGCAATCGCGCTCCACCCCGAGGGGTACCTCATACCCCGGG 501
QY 490 aagttcatcaacatgttgcctgtgagtgagctaaagaacatgaagtcagcgtaag 549
Db 502 TGGCTTTGGGAGAGTACGGCTGAGCGCGGACGAGTGCGGTGAGTACCGGGCCCTGGCG 561
QY 550 aatgacactccctacaaagaggcagacagttctctcggaagatgagatccccag 609
Db 562 GGGGACCCCTCCGACAACTCCCGGGGTCAAGGGCATCGGGGAGAGACCGCCNGAAG 621
QY 610 tctatcagagtcgcagacactttcatgttctctgcccacacacacagatcaccccag 669
Db 622 CTCTCNAGAGTGGGGAGCTTGGAAACCTCTCAAGAACTGGACCGGGTGAAGCCC 681
QY 670 tgtctccaccagcaacttgaag 691
Db 682 GCCNTCCGGGAGAGATCCAGG 703

RESULT 12
US-08-757-653-162
/ Sequence 162, Application US/08757653
/ Patent No. 5843669
/ GENERAL INFORMATION:
/ APPLICANT: Kaiser, Michael W.
/ APPLICANT: Lyamichiev, Victor I.
/ APPLICANT: Lyamichiev, Natasha
/ TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
/ TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
/ NUMBER OF SEQUENCES: 190
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Medlen & Carroll, LLP
/ STREET: 220 Montgomery Street, Suite 2200
/ CITY: San Francisco
/ STATE: California
/ COUNTRY: United States Of America
/ ZIP: 94104
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/757,653
/ FILING DATE:

/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ingolia, Diane E.
/ REGISTRATION NUMBER: 40,027
/ REFERENCE/DOCKET NUMBER: FORS-02565
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 705-8410
/ TELEFAX: (415) 397-8338
/ INFORMATION FOR SEQ ID NO: 162:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 963 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..960
/ US-08-757-653-162

Query Match 4.3%; Score 38; DB 3; Length 963;
Best Local Similarity 45.4%; Pred. No. 0.19;
Matches 137; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 430 ctgtgcacatccgcagcaggaagactttgtctctgagccctacccctgacctgtgc 489
Db 493 CAGCTCTTTCCGACCGCAATCCACGTCTCCACCCCGAGGGGTACCTCATACCCCGGG 552
QY 490 aagttcatcaacatgttgcctgtgagtgagctaaagaacatgaagtcagcgtaag 549
Db 553 TGGCTTTGGGAGAGTACGGCTGAGCGCGGACGAGTGCGGTGAGTACCGGGCCCTGGACC 612
QY 550 aatgacactccctacaaagaggcagacagttctctcggaagatgagatccccag 609
Db 613 GGGGACCGAGTCCGACAACTCCCGGGGTCAAGGGCATCGGGGAGAGACCGCGGGAAG 672
QY 610 tctatcagagtcgcagacactttcatgttctctgcccacacacacagatcaccccag 669
Db 673 CTCTCGAGAGTGGGGAGCTTGGAAACCTCTCAAGAACTGGACCGGGTGAAGCCC 732
QY 670 tgtctccaccagcaacttgaagtcacccaggtctatgagagctgtggtgacattgtc 729
Db 733 GCCATCCGGGAGAGATCCTGCGCCCATGACGATCTGAAGCTCTCTGGGACCTGGCC 792
QY 730 aa 731
Db 793 AA 794

RESULT 13
US-08-823-516-60
/ Sequence 60, Application US/08823516
/ Patent No. 5994069
/ GENERAL INFORMATION:
/ APPLICANT: Hall, Jeff G.
/ APPLICANT: Lyamichiev, Victor I.
/ APPLICANT: Mast, Andrea L.
/ APPLICANT: Brow, Mary Ann D.
/ TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
/ TITLE OF INVENTION: Sequential Invasive Cleavages
/ NUMBER OF SEQUENCES: 163
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Medlen & Carroll, LLP
/ STREET: 220 Montgomery Street, Suite 2200
/ CITY: San Francisco
/ STATE: California
/ COUNTRY: United States Of America
/ ZIP: 94104
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS

```

1 SOFTWARE: PatentIn Release #1.0, Version #1.30
2
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/823,516
5 FILING DATE: 24-MAR-1997
6 CLASSIFICATION: 435
7
8 PRIOR APPLICATION DATA:
9 APPLICATION NUMBER: PCT/US97/01072
10 FILING DATE: 21-JAN-1997
11
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: US 08/759,038
14 FILING DATE: 02-DEC-1996
15
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 08/758,314
18 FILING DATE: 02-DEC-1996
19
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 08/756,386
22 FILING DATE: 29-NOV-1996
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 08/682,853
26 FILING DATE: 12-JUL-1996
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 08/599,491
30 FILING DATE: 24-JAN-1996
31
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Ingollia, Diane E.
34 REGISTRATION NUMBER: 40,027
35 REFERENCE/DOCKET NUMBER: FORS-02736
36
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: (415) 705-8410
39 TELEFAX: (415) 397-8338
40
41 INFORMATION FOR SEQ ID NO: 60:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 963 base pairs
44 TYPE: nucleic acid
45 STRANDEDNESS: double
46 TOPOLOGY: linear
47 MOLECULE TYPE: DNA (genomic)
48
49 FEATURE:
50 NAME/KEY: CDS
51 LOCATION: 1..960
52
53 US-08-823-516-60

```

Query Match	4.3%	Score 38;	DB 4;	Length 963;	
Best Local Similarity	45.4%;	Pred. No. 0.19;			
Matches 137;	Conservative	0;	Mismatches 165;	Indels 0;	Gaps
Qy	430	ctgtgccatgccagagcgaggaagaactttgtctctgagccctactctctgaccttggc	489		
Db	493	CAGCTCCTTTCCAGCCGCATCCACGTCTCCACCCGAGGGGTACCTCATCACCCGGCC	552		
Qy	490	aagttcatcaacatgtttgctgctctggtatgagctaaagaacatgaagtgcagcgtcaag	549		
Db	553	TGGCTTTGGGAAAAGTACGGCTTGAGGCCCGACACGTGGCGCCACTACCGGGCCCTGACC	612		
Qy	550	aatgaccactccgcctacaagagggcagcacagttctctcggaagatggcagatccccag	609		
Db	613	GGGGACGAGTCCGACAACTTCCCGGGGTCAAGGGCATCGGGGAGAAACACGGCGAGGAAG	672		
Qy	610	tctatccaggagtgcgcaacctttccatgtttctctggcaaccacacaaaggatcacccag	669		
Db	673	CTTCTGGAGGAGTGGGGGAGCCCTGGAAGCCCTCCTCAAGAACCTGGACCGGTGGAAGCC	732		
Qy	670	tgtctccaccagcaacttgaagtgatccccaggctatggaggagtctggtgcacatgttc	729		
Db	733	GCCATCCGGGAGAGATCTCTGGGCCACATGGACGATCTGAAGCTCTCTCTGGGACCTGGCC	792		
Qy	730	aa	731		
Db	793	aa	794		

RESULT 14

RESULT 14

```

US-08-073-384C-12
; Sequence 12, Application US/08073384C
; Patent No. 5541311
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,384C
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-073-384C-12

```

Query Match	4.3%	Score 38;	DB 1;	Length 1600;
Best Local Similarity	45.4%;	Pred. No. 0.23;		
Matches 137;	Conservative 0;	Mismatches 165;	Indels 0;	Gaps 0;

Qy	430	ctgtgcaatgcgcagcaggagaagtatttctctgagccactacctctgaaccccttgcc	489
Db	443	CAGTCCTCTTCCGACCGCATCAGCTCTCACCCCAGGGGTACTCTATCAACCCCGGCC	502
Qy	490	aagttcatcaaatgtttgctgtccctgatgagctaagaagaaacatgaagtgcagcgctcaag	549
Db	503	TGGCTTTGGGAAGAAGTACGGCTGAGGCCCGACCAGTAGTGCGCGCACTACCGGGCCCTGACC	562
Qy	550	aatgaccactcgcctacaagaaggcgacagattcctcgggaaagatggcagatccccag	609
Db	563	GGGGACGAGTCCGCACAACCTTCCCGGGGTCAAGGGCATCGGGGGAAGACGGCGAGGAAG	622
Qy	610	tctatccaggagtcgcagaacctttcccattgctctggccaaccaaacaggatcacccag	669
Db	623	CTTCTGGAGGAGTGGGGGAGCCTGGAA GCCCTTCCTCAAGAACCTGGACCGGCTGAAGCC	682
Qy	670	tgtctccaccagcaacttgaagtgatcccgagctatgaggagagctgctggctgacattgtc	729
Db	683	GCCATCCGGGAGAAAGATCCTGGCCCCCATGGACGATCTGAAGCTCTCTCTGGGACCTGGCC	742
Qy	730	aa 731	
Db	743	AA 744	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:01:08 ; Search time 3550.87 Seconds
(without alignments)
-1048.473 Million cell updates/sec

Title: US-09-215-435-113
Perfect score: 893
Sequence: 1 gccatgatcacgcagctcac.....aaaaaaaaaaaaaaaaaaaa 893

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues

Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA:*

1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/1/pna/US08_COMB.seq:*

5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq:*

6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq:*

7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq:*

8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq:*

9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq:*

10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq:*

11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq:*

12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq:*

13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq:*

14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq:*

15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq:*

16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq:*

17: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*

18: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*

19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq:*

20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq:*

21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq:*

22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq:*

23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq:*

24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq:*

25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq:*

26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq:*

27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq:*

28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq:*

29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq:*

30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq:*

31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq:*

32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq:*

33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq:*

34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq:*

35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq:*

36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq:*

37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq:*

38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq:*

39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq:*

40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq:*

41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq:*

42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq:*

43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq:*

44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq:*

45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq:*

46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq:*

47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq:*

48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq:*

49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq:*

50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq:*

51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq:*

52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq:*

53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq:*

54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq:*

55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq:*

56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq:*

57: /cgn2_6/ptodata/1/pna/US095_COMB.seq:*

58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*

59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq:*

60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq:*

61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq:*

62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq:*

63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq:*

64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq:*

65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq:*

66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq:*

67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq:*

68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq:*

69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*

70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*

71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*

72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq:*

73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq:*

74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq:*

75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq:*

76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq:*

77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq:*

78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq:*

79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq:*

80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq:*

81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq:*

82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq:*

83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq:*

84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq:*

85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq:*

86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq:*

87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq:*

88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*

89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq:*

90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq:*

91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq:*

92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq:*

93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq:*

94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq:*

95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq:*

96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq:*

97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq:*

98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*

99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*

100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*

101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*

102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*

103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*

104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	893	100.0	893	43	US-09-215-435-113 Sequence 113, App
2	872	97.6	1315	43	US-09-215-435-260 Sequence 260, App
3	872	97.6	1315	70	US-60-069-957-93 Sequence 93, Appl

```

4 775.2 86.8 1734 51 US-09-354-899-7143 Sequence 7143, Ap
5 775.2 86.8 1734 52 US-09-371-168-6632 Sequence 6632, Ap
6 714.2 80.0 6569 92 US-09-172-360-21713 Sequence 21713, A
7 664.2 74.4 878 56 US-09-471-275-3565 Sequence 3565, Ap
8 603.8 67.6 4403 51 US-09-359-922-6045 Sequence 6045, Ap
9 603.8 67.6 4403 51 US-09-359-922-6045 Sequence 6045, Ap
10 571.8 64.0 4400 43 US-09-205-070-12698 Sequence 12698, A
11 571.8 64.0 4400 50 US-09-340-623-12698 Sequence 12698, A
12 571.8 64.0 4474 56 US-09-471-275-1734 Sequence 1734, Ap
13 571.8 64.0 4823 92 US-09-172-360-19789 Sequence 19789, A
14 456 51.1 1827 45 US-09-270-767-14852 Sequence 14852, A
15 456 51.1 3986 90 US-09-167-217-10144 Sequence 10144, A
16 456 51.1 3986 92 US-09-171-627-860 Sequence 860, App
17 456 51.1 3986 93 US-09-173-464-8178 Sequence 8178, Ap
18 397 44.5 444 53 US-09-399-720-16879 Sequence 16879, A
19 376 42.1 6458 90 US-09-167-217-10143 Sequence 10143, A
20 376 42.1 6458 92 US-09-171-627-859 Sequence 859, App
21 376 42.1 6458 93 US-09-173-464-8177 Sequence 8177, Ap
22 376 42.1 28952 88 US-09-150-584-168 Sequence 168, App
23 376 42.1 28952 88 US-09-150-584-432 Sequence 432, App
24 364.8 40.9 463 45 US-09-287-618-9033 Sequence 9033, Ap
25 320.8 35.9 3548 93 US-09-173-464-10923 Sequence 10923, A
26 320.8 35.9 3551 90 US-09-167-217-13406 Sequence 13406, A
27 318.2 35.6 447 46 US-09-287-618-26441 Sequence 26441, A
28 305.2 34.2 493 47 US-09-293-972-21120 Sequence 21120, A
29 304 34.0 426 56 US-09-496-911-13386 Sequence 13386, A
30 298.8 33.5 397 52 US-09-371-168-15000 Sequence 1500, Ap
31 280.4 31.4 418 51 US-09-359-067-47954 Sequence 47954, A
32 239.6 25.8 407 56 US-09-496-911-13388 Sequence 13388, A
33 228.4 25.6 366 8 US-09-196-363A-4580 Sequence 4580, Ap
34 228.4 25.6 366 8 US-09-196-363A-4580 Sequence 4580, Ap
35 228.4 25.6 366 8 US-09-196-363-4580 Sequence 4580, Ap
36 222.8 24.9 279 38 US-09-070-694-5511 Sequence 5511, Ap
37 222.8 24.9 307 67 US-09-044-030-3917 Sequence 3917, Ap
38 220.4 24.7 307 20 US-09-766-606-823 Sequence 823, App
39 220.4 24.7 307 32 US-09-979-854A-823 Sequence 823, App
40 217.2 24.3 351 50 US-09-342-216-819 Sequence 819, App
41 212.4 23.8 290 38 US-09-073-079-1927 Sequence 1927, Ap
42 212.4 23.8 290 68 US-09-047-802-1927 Sequence 1927, Ap
43 208.4 23.3 400 46 US-09-287-618-2702 Sequence 2702, Ap
44 197.4 22.1 331 27 US-09-918-972-1073 Sequence 1073, Ap
45 197.4 22.1 331 62 US-09-025-133-1073 Sequence 1073, Ap
```

ALIGNMENTS

```

RESULT 1
US-09-215-435-113
; Sequence 113, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 113
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```

; NAME/KEY: CDS
; LOCATION: 4..810
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 4..279
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 6.8
; OTHER INFORMATION: seq AVMLYTWRSRA/IP
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 858..863
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 881..893
; US-09-215-435-113

Query Match 100.0%; Score 893; DB 43; Length 893;
Best Local Similarity 100.0%; Pred. NO. 1.8e-198;
Matches 893; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccatgacgcagcgtcacccctggaagatgccctgtccaaacgtggacctgtctgaagag 60
Db 1 gccatgacgcagcgtcacccctggaagatgccctgtccaaacgtggacctgtctgaagag 60
Qy 61 ctccccctccccaccagcagccatgcacgcagcctccaccccttccatccatcatgaccag 120
Db 61 ctccccctccccaccagcagccatgcacgcagcctccaccccttccatccatcatgaccag 120
Qy 121 gctaacttggacaaactttggagacaggaatgcatttgcacggggcattgcgaaggtac 180
Db 121 gctaacttggacaaactttggagacaggaatgcatttgcacggggcattgcgaaggtac 180
Qy 181 attgacgaggtacagtcacccatccacgcagcagcagcagcagcagcagcagcagcag 240
Db 181 attgacgaggtacagtcacccatccacgcagcagcagcagcagcagcagcagcagcag 240
Qy 241 gcggctcagctgtcacacctggcgagcgtgttcccccggggccattccccaggtgaaatgc 300
Db 241 gcggctcagctgtcacacctggcgagcgtgttcccccggggccattccccaggtgaaatgc 300
Qy 301 gagcagcccaacccaggtagagatctatgagacagcagcagcagcagcagcagcagcag 360
Db 301 gagcagcccaacccaggtagagatctatgagacagcagcagcagcagcagcagcagcag 360
Qy 361 accaagctcatgaagtctcatgtattttcagcgcagcagcagcagcagcagcagcagcag 420
Db 361 accaagctcatgaagtctcatgtattttcagcgcagcagcagcagcagcagcagcagcag 420
Qy 421 gtgaagcggctgtgcccagcgcagcagcagcagcagcagcagcagcagcagcagcagcag 480
Db 421 gtgaagcggctgtgcccagcgcagcagcagcagcagcagcagcagcagcagcagcagcag 480
Qy 481 acccttggcaagttcatcaacatgtttgctgctcctggaggaagcagcagcagcagcagcag 540
Db 481 acccttggcaagttcatcaacatgtttgctgctcctggaggaagcagcagcagcagcagcag 540
Qy 541 agcgtcaagaatgaccactccgcctacaaaggggagcagcagcagcagcagcagcagcagcag 600
Db 541 agcgtcaagaatgaccactccgcctacaaaggggagcagcagcagcagcagcagcagcagcag 600
Qy 601 gatccccagtcctaccaggtcgagaaacctttccatgttctctggcccaaccacacaggg 660
Db 601 gatccccagtcctaccaggtcgagaaacctttccatgttctctggcccaaccacacaggg 660
Qy 661 atcacccagtcctccacagcaacttgaagtgtatccaggcagcagcagcagcagcagcagcag 720
Db 661 atcacccagtcctccacagcaacttgaagtgtatccaggcagcagcagcagcagcagcagcag 720
Qy 721 gacattgtcaacatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
Db 721 gacattgtcaacatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
```


TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 1315 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: CDS
LOCATION: 123..302
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 123..176
IDENTIFICATION METHOD: von Heijne matrix
OTHER INFORMATION: score 4.3
OTHER INFORMATION: seq WTLKSPPTSS/HA
FEATURE:
NAME/KEY: poly_a_signal
LOCATION: 1279..1284
FEATURE:
NAME/KEY: poly_a
LOCATION: 1301..1312
IDENTIFICATION METHOD: blastn
FEATURE:
NAME/KEY: est
LOCATION: 628..868
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: Identity 95
OTHER INFORMATION: region 1..241
OTHER INFORMATION: id AA143673
US-60-069-957-93

Query Match 97.6%; Score 872; DB 70; Length 1315;
Best Local Similarity 98.2%; Pred. No. 1.6e-193;
Matches 866; Conservative 13; Mismatches 3; Indels 0; Gaps 0;
QY 1 gccatgacacgcacgtcacctcgaagatgacctgtgtccaaagtggacctgttgaaag 60
DB 95 GCATGACCACGACGTCACCTGGAAGATGCCCTGTCCAAGTGGACCTGCTGAAGAG 154
QY 61 cttccctcccgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 120
DB 155 CTTCCCTCCCCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 214
QY 121 gctaaatttgacacaaacttgagacaggaatgcatgttgacgggcatgtgcaaggtac 180
DB 215 GCTAACTTTGACACAACTTTGAGGACAGGANTGATTTGTTCACGGGCGATTGCAAGGTAC 274
QY 181 attgagcaggtacagtcacctccagcatgaatgagatgctggaggaaggacatgatat 240
DB 275 ATTGACGAGGTACAGTCCATCCAGCATGATGATGCTGGAGGAGGACATGAKTAT 334
QY 241 gcggtatgctgtaacacctggcgaactgttccggggccattcccaggtgaaatgcaac 300
DB 335 GCGGTATGCTGTACACCTGGCGACGCTGTCCGGGGCCATTCCTCCAGGTGAATGCAAC 394
QY 301 gacgagcccaacgagtagagatctatgagaagacagtagagtgctggagccgaggtc 360
DB 395 RACGAGCCCAACCGAKTARARATCTATGARAARACAGTARAGTGTGTGGAGCGGAGTTC 454
QY 361 accaagctaatgaagtctatgtatttttcagcgaaggccatgagcggttctgcagcag 420
DB 455 ACCAAGCTCATGAAGTTCATGTTATTTTCACGCAAGGCCATCGAGGGGTCTGCAACGAG 514
QY 421 gtaagcggctgtgccatgcccgcgcaggaaggaactttgtctctgagccctacctctcg 480
DB 515 GTGAAGCGGCTGTGCCATGCCGACGCGCAGGAGGACTTTGTCTCTGAGGCGCTACCTCCTG 574

QY 481 acccttgccaagtctcatcaacatgttctgtctctgtagtgagctaaagacatgaagtgc 540
DB 575 ACCCTTGGCAAGTTTATCAACATGTTGTGCTCTCTGGATGAGCTAAAGAACATGAATGC 634
QY 541 agcgtcaagaatgaccactccgctacaaagaggggacagtcctctcggaagatggca 600
DB 635 AGCGTCAARAATGATGACCACTCTGCTTACAAGAGGGGACAGTTCCTCGGAAGATGGCA 694
QY 601 gatccccagtctatccagagtcgagaaacaccttccatgttctggcccaaccacacag 660
DB 695 GATCCCCAGTCTATCCAGGAGTGCAGAACCTTTCCATGTTCTGTGCCCACACACAGG 754
QY 661 atcacccagtgtctccaccagcaacttgaaagtatccccaggctatgagagctgctggct 720
DB 755 ATCACCAGTGCTCCACCAGCAACTTGAAGTATCCAGGCTATGAGGAGCTGCTGGCT 814
QY 721 gacattgtcaacatctgtgtgattactacgagaaacaagatgtacatgactcccaatgag 780
DB 815 GACATTGTCAACATCTGTGTGGATTACTACGAGAACAAGATGTACCTGACTCCCACTGAG 874
QY 781 aaacatatgctcctcaaggtaaaactccccctgagcgccaccatgagcctggcgttac 840
DB 875 AAACATATGCTCTCAAGGTAAACCTCCCTGAGCGCGACCCATGGAGCCTGGGCTTAC 934
QY 841 cctctcaccttcttcttattataaaatccggttttaaaaaacaa 882
DB 935 CCTCTCACCTTCTTCTTATTAAAAATCCGTTTAAAAAACAA 976

RESULT 4
US-09-354-899-7143
; Sequence 7143, Application US/09354899
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, M. Alexandra
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
; FILE REFERENCE: MLN98-35Pa
; CURRENT APPLICATION NUMBER: US/09/354,899
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: 60/092,406
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 8631
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7143
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1734)
; OTHER INFORMATION: n = A,T,C or G
US-09-354-899-7143
Query Match 86.8%; Score 775.2; DB 51; Length 1734;
Best Local Similarity 98.8%; Pred. No. 6.8e-171;
Matches 792; Conservative 0; Mismatches 8; Indels 2; Gaps 1;
QY 1 gccatgatcacgcacgtcacctcgaagatgacctgtccaaagtggacctgttgaaag 60
DB 152 gccatgaccacgcacgtcacctcgaagatgacctgtccaaagtggacctgttgaaag 211
QY 61 cttccctcccgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 120
DB 212 cttccctcccgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 269
QY 121 gctaaatttgacacaaacttgagacaggaatgcatgttcacggggcattgcaaggtac 180
DB 270 gctaaatttgacacaaacttgagacaggaatgcatgttcacggggcattgcaaggtac 329
QY 181 attgagcaggtctacagtcacctccagcatgaatgagatgctggaggaaggacatgatat 240


```
Db 330 attgagcaggctacagctccactccagcatgaatgagatgctgagggaagacatgagtat 389
QY 241 gcggtcatgcttacacctgcgcagctgtcccgagccattcccccaggtgaaatcaac 300
Db 390 gcggtcatgcttacacctgcgcagctgtcccgagccattcccccaggtgaaatcaac 449
QY 301 gacagcccaaccgagtagagatctatgagaacacagtagagtgctggagccggaggtc 360
Db 450 gacagcccaaccgagtagagatctatgagaacacagtagagtgctggagccggaggtc 509
QY 361 accaagctcatgaagtcatgtatttttcagcgcaaggccaatcgagcggttcttcagcag 420
Db 510 accaagctcatgaagtcatgtatttttcagcgcaaggccaatcgagcggttcttcagcag 569
QY 421 gtgaagcggctgtgcacatgcgcagcgcaaggaaactttgtctctgagccctaccctcg 480
Db 570 gtgaagcggctgtgcacatgcgcagcgcaaggaaactttgtctctgagccctaccctcg 629
QY 481 acccttggaagttcatcaacatgtttgtctctgctgtagagctaaagaacatgaagtgc 540
Db 630 acccttggaagttcatcaacatgtttgtctctgctgtagagctaaagaacatgaagtgc 689
QY 541 agcgtcaagaatgacacactcgcctcaacagagggcgacacagttcctgcggaagatggca 600
Db 690 agcgtcaagaatgacacactcgcctcaacagagggcgacacagttcctgcggaagatggca 749
QY 601 gatcccaagtctacagagtcgcagaaacctttccatgttcttgcccaaccaacagg 660
Db 750 gatcccaagtctacagagtcgcagaaacctttccatgttcttgcccaaccaacagg 809
QY 661 atcaccagtgctccaccagcaacttgaagtgtatcccaggctatgaggagctgctggct 720
Db 810 atcaccagtgctccaccagcaacttgaagtgtatcccaggctatgaggagctgctggct 869
QY 721 gacattgtcaacatctgtgtgattactacgagaacaaagtgtacctgactcccaagttag 780
Db 870 gacattgtcaacatctgtgtgattactacgagaacaaagtgtacctgactcccaagttag 929
QY 781 aaacatatgctctccaagtaa 802
Db 930 aaacatatgctctccaagtgta 951
```

RESULT 5

```
US-09-371-168-6632
; Sequence 6632, Application US/09371168
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, M. Alexandra
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
; FILE REFERENCE: MLN98-37pA
; CURRENT APPLICATION NUMBER: US/09/371,168
; EARLIER FILING DATE: 1999-08-10
; EARLIER FILING DATE: 1998-08-10
; EARLIER FILING DATE: 1998-08-10
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 8285
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6632
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1734)
; OTHER INFORMATION: n = A,T,C or G
US-09-371-168-6632
```

Query Match 86.8%; Score 775.2; DB 52; Length 1734;
Best Local Similarity 98.8%; Pred. No. 6.8e-171;
Matches 792; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

```
QY 1 gccatgatcacgcacgtcaccttggaagatgcctgtccaaacgtggacctgcttgaagag 60
Db 152 gccatgaccacgcacgtcaccttggaagatgcctgtccaaacgtggacctgcttgaagag 211
QY 61 ctccctcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 120
Db 212 ctccctcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 269
QY 121 gctaaatttgacacaaactttgaggacaggaatgcattttgttcacgggcatctgcaaggtac 180
Db 270 gctaaatttgacacaaactttgaggacaggaatgcattttgttcacgggcatctgcaaggtac 329
QY 181 attgagcaggctacagtccactccagcatgaatgagatgctgtaggaagacatgagtat 240
Db 330 attgagcaggctacagtccactccagcatgaatgagatgctgtaggaagacatgagtat 389
QY 241 gcggtcatgctgtacacctggcgagctgttcccgggccaattcccaggtgaaatcaac 300
Db 390 gcggtcatgctgtacacctggcgagctgttcccgggccaattcccaggtgaaatcaac 449
QY 301 gacagcccaaccgagtagagatctatgagaacacagtagagtgctggagccggaggtc 360
Db 450 gacagcccaaccgagtagagatctatgagaacacagtagagtgctggagccggaggtc 509
QY 361 accaagctcatgaagtcatgtatttttcagcgcaaggccaatcgagcggttcttcagcag 420
Db 510 accaagctcatgaagtcatgtatttttcagcgcaaggccaatcgagcggttcttcagcag 569
QY 421 gtgaagcggctgtgcacatgcgcagcgcaaggaaactttgtctctgagccctaccctcg 480
Db 570 gtgaagcggctgtgcacatgcgcagcgcaaggaaactttgtctctgagccctaccctcg 629
QY 481 acccttggaagttcatcaacatgtttgtctctgctgtagagctaaagaacatgaagtgc 540
Db 630 acccttggaagttcatcaacatgtttgtctctgctgtagagctaaagaacatgaagtgc 689
QY 541 agcgtcaagaatgacacactcgcctcaacagagggcgacacagttcctgcggaagatggca 600
Db 690 agcgtcaagaatgacacactcgcctcaacagagggcgacacagttcctgcggaagatggca 749
QY 601 gatcccaagtctacagagtcgcagaaacctttccatgttcttgcccaaccaacagg 660
Db 750 gatcccaagtctacagagtcgcagaaacctttccatgttcttgcccaaccaacagg 809
QY 661 atcaccagtgctccaccagcaacttgaagtgtatcccaggctatgaggagctgctggct 720
Db 810 atcaccagtgctccaccagcaacttgaagtgtatcccaggctatgaggagctgctggct 869
QY 721 gacattgtcaacatctgtgtgattactacgagaacaaagtgtacctgactcccaagttag 780
Db 870 gacattgtcaacatctgtgtgattactacgagaacaaagtgtacctgactcccaagttag 929
QY 781 aaacatatgctctccaagtaa 802
Db 930 aaacatatgctctccaagtgta 951
```

RESULT 6

```
US-60-172-360-21713
; Sequence 21713, Application US/60172360
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE REFERENCE: Polynucleotide Sequence Databases, and Single Nucleotide Polym
; CURRENT APPLICATION NUMBER: US/60/172,360
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 29838
; SOFTWARE: PERL Program
; SEQ ID NO 21713
```



```
Db 256 attgagcaggtacagtcactccactccagcatgaatgagatgctggagggaagacatgagtat 315
QY 241 gcggtcatgtgtacacgtgcgcagctgttcccgccattcccccaggtgaaatgaac 300
Db 316 gcggtcatgtgtacacgtgcgcagctgttcccgccattcccccaggtgaaatgaac 375
QY 301 gaggagcccaacccagtagagatctatgagaagacagtagagtgctggagccggaggtc 360
Db 376 gaggagcccaacccagtagagatctatgagaagacagtagagtgctggagccggaggtc 435
QY 361 accaagctcatgaattcatgtatttttcacgcaaggccaatgcagcggtcttcgcagcgag 420
Db 436 accaagctcatgaattcatgtatttttcacgcaaggccaatgcagcggtcttcgcagcgag 495
QY 421 gtaagcggctgtgccatgcgcagcgaggaagacctttctctgagcgctaccctctg 480
Db 496 gtaagcggctgtgccatgcgcagcgaggaagacctttctctgagcgctaccctctg 555
QY 481 accttggaagttcatcaacatgtttctgtctctgagtagagtaagaacatgaagtgc 540
Db 556 accttggaagttcatcaacatgtttctgtctctgagtagagtaagaacatgaagtgc 615
QY 541 agctcaagaataaccactccgctcaacaagagggcagcacagttctctgcgaagatggca 600
Db 616 agctcaagaataaccactccgctcaacaagagggcagcacagttctctgcgaagatggca 675
QY 601 gatccccagttctccagaggtgcgagaacctttccatgttctctgcccacacacagc 660
Db 676 gatccccagttctccagaggtgcgagaacctttccatgttctctgcccacacacagc 735
QY 661 atcacccag 669
Db 736 atcacccag 744

RESULT 8
US-09-359-922-6045/c
; Sequence 6045, Application US/09359922
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922
; EARLIER FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6045
; LENGTH: 4403
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-6045

Query Match 67.6%; Score 603.8; DB 51; Length 4403;
Best Local Similarity 84.7%; Pred. No. 8.6e-131;
Matches 677; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 4 atgatacgcacgtccactccagcatgaatgagatgctggagggaagacatgagtatgcy 243
```

```
Db 4168 GAACAACCCACCGTCCACTCTAGCATCAACGAGATGCTGGAGGAGGCCAGAAATATGCT 4109
QY 244 gtcagtgttacacctggcgcagctgttcccgccattcccccaggtgaaatgaacgag 303
Db 4108 GTCATGTGTACACCTGGAGAGCTGTCCCGGCCATCCACAGGTGAAATGTAACGAG 4049
QY 304 cagcccaacccagtagagatctatgagaagacagtagagtgctggagccggaggtcacc 363
Db 4048 CAGCCTAACAGAGTGGAAATCTACGAGAAACCGTGGAGGTTCTGGAGCCCTACGTCACA 3989
QY 364 aagctcatgaagttcatgtatttttcacgcaaggccaatgcagcggttcttcgcagcgagtg 423
Db 3988 AAATGATGAATTTTCATGTACTTCCAGAGAAATGCCATTGAGCGTTTCTCGGGGAAGTG 3929
QY 424 aagcggctgtgcacatgccgagcaggaagacctttctctgagcgctaccctctgacc 483
Db 3928 AGCGCGCTGTGCCATGCCGAGGAGGAGGACTTCTGTGTGAGAAGCTTACCTGATCACA 3869
QY 484 ctggcaagttcatcaacatgtttctgtctctgagtagagctaaagaacatgaagtgcagt 543
Db 3868 CTGGGCAAAATTCACATGTTGCTGTGCTGTGACGAGCTGAAGAACATGAAGTGCAGT 3809
QY 544 gtcagaatgaccactccgctcaacaagggcagcacagttctctgcggaagatggcgagat 603
Db 3808 GTGAAGAACACACCTCTGCTTACAAGAGGGCAGCACAGTTCTTCCGGAAGATGGCAGAT 3749
QY 604 cccagctctacagagtgctgcgaacctttccatgttctctgcccacacacacaggtc 663
Db 3748 CCCAGTCTCCATCCAGGAGTGCAGAACCTTTTCCATGTTCTTGGCCAAACACACAGGATC 3689
QY 664 accagtgctccacagcaacttgaagtatcccgagctatcccgagctatgagagctgctggctgac 723
Db 3688 ACCAGTATCTCCACAGCACTTGAAGTATTTCTGGCTACGGAAGAGCTTCTTGGCAGAT 3629
QY 724 attgtcaacatctgtgtggattactacgagaacaaagatgtacctgtacctccagtgagaaa 783
Db 3628 ATTGTGAATCTGTGTGGATTACTACGAGAACAGGATGTATTTGAGCCCCCAGTGAGAAA 3569
QY 784 catatgctctcacaaggtaa 802
Db 3568 CACATGCTTCTCAAAAGTCA 3550

RESULT 9
US-09-359-922-6045/c
; Sequence 6045, Application US/09359922A
; GENERAL INFORMATION:
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Liu, Jin
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922A
; EARLIER FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: US 09/034,341
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6045
; LENGTH: 4403
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-6045

Query Match 67.6%; Score 603.8; DB 51; Length 4403;
Best Local Similarity 84.7%; Pred. No. 8.6e-131;
Matches 677; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 4 atgatacgcacgtccactccagcatgaatgagatgctcccaagtgacacctgttgaagagctt 63
```


EARLIER APPLICATION NUMBER: US 09/205,070
EARLIER FILING DATE: 1998-12-03
NUMBER OF SEQ ID NOS: 45207
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12698
LENGTH: 4400
TYPE: DNA
ORGANISM: Homo sapiens
US-09-340-623-12698

Query Match 64.0%; Score 571.8; DB 50; Length 4400;
Best Local Similarity 82.2%; Pred. No. 2.5e-123;
Matches 657; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 4 atgatacagcaagtcacccctggaagatgccctgtccaaactgagacactgttgaagagctt 63
DB 56 atggcgccaggtgactctgagagcgctgtccaaactgagacactgtgagagctg 115
QY 64 cccctcccgacagcagcatgcatgagcctccaccttctccatcatgtaccagct 123
DB 116 cccctgcccagcagcagccctgcatgagcccgccatctctgctctaccagcca 175
QY 124 aacttgacacaaacttgagacaggaatgcatttgtcacgggcatgtcaaggtacatt 183
DB 176 aatttcaacactaaacttgagacagaaatgcatttgttactggcatgcgaagatacatt 235
QY 184 gaggcagctacagtcaccctccagcatgaatgagatgctgagagaagacatgagtgcg 243
DB 236 gaacagccaccctccactctagcatgaacagatgctgagaggggccaagaatgct 295
QY 244 gtcatgtgtacacctggcgagctgttcccgccattcccccaggtgaaatgcaacgag 303
DB 296 gtcatgtgtacacctggagagctgtctcccgccatcccccaggtgaaatgcaacgag 355
QY 304 cagcccaacccagtagtagatctatgagaagacagtagaggtgtgctggagccggaggtcacc 363
DB 356 cagcctaacagagtgaaatctacgagaaacccgtgaggtttctggagcctgaggtcacca 415
QY 364 aagctcatgaattcatgtattttcagcgaagggccatcagcaggtttctgacgaggtg 423
DB 416 aaactgatgaattcatgtactctcagagaaatgccattgagcttctgctggggaagtg 475
QY 424 aagcgctgtgcatgccgagcgaggaagactttgtctgagccctacctcctgacc 483
DB 476 aagcgctgtgcatgccgagaggaaggaagacttctgtctgagagcctacctgatacaca 535
QY 484 ctggcaagttcatcaacatgtttgtctgtcctggatgagctaaagaacatgaagtgcagc 543
DB 536 ctgggcaaatctcaacatgttctgtgtgctggagagctgaagaacatgaagtgcagt 595
QY 544 gtcaagaatgacactccgcttacaagaagggcagacagcttctgaggagatggcagat 603
DB 596 gtgaagaacgacactcagctgacagagggcgctcagtttttacgtaaaaatggcagat 655
QY 604 cccagctctatcagagtcgcagaaacctttcatgttcttctgagccaaacacacagagtc 663
DB 656 ccacagctccatccaggaaatcgagaatctgtccatgttcttctgagccatatacaagaatc 715
QY 664 accagtgctccaccagcaacttgaagtgtatcccgaggtatgagagatgctgtgctgac 723
DB 716 acacagctctcagcagcagctcgaagtgtatttctgctagagaagctcctctggcagat 775
QY 724 attgtcaacatctgtgtgattactacgagaacagatgtacctgactcccgatgagaaa 783
DB 776 attgtgaactctgtgtgattactacgagaacagatgtattttgacgcccagtgagaaa 835
QY 784 catatgctcctcaagtaa 802
DB 836 cacatgcttctcaagta 854

RESULT 12

US-09-471-275-1734
Sequence 1734, Application US/09471275
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: Novel Contigs Obtained
TITLE OF INVENTION: From Various Libraries
FILE REFERENCE: 782
CURRENT APPLICATION NUMBER: US/09/471,275
CURRENT FILING DATE: 1999-12-23
EARLIER APPLICATION NUMBER: US 09/235,076
EARLIER FILING DATE: 1999-01-20
EARLIER APPLICATION NUMBER: US 09/234,611
EARLIER FILING DATE: 1999-01-22
EARLIER APPLICATION NUMBER: US 09/240,371
EARLIER FILING DATE: 1999-01-29
EARLIER APPLICATION NUMBER: US 09/277,227
EARLIER FILING DATE: 1999-03-25
EARLIER APPLICATION NUMBER: US 09/271,490
EARLIER FILING DATE: 1999-03-18
EARLIER APPLICATION NUMBER: US 09/293,972
EARLIER FILING DATE: 1999-04-15
EARLIER APPLICATION NUMBER: US 09/274,861
EARLIER FILING DATE: 1999-03-23
EARLIER APPLICATION NUMBER: US 60/125,453
EARLIER FILING DATE: 1999-03-19
EARLIER APPLICATION NUMBER: US 60/126,605
EARLIER FILING DATE: 1999-03-26
EARLIER APPLICATION NUMBER: US 09/306,350
EARLIER FILING DATE: 1999-05-07
EARLIER APPLICATION NUMBER: US 09/399,720
EARLIER FILING DATE: 1999-09-21
EARLIER APPLICATION NUMBER: US 09/404,284
EARLIER FILING DATE: 1999-09-21
EARLIER APPLICATION NUMBER:
EARLIER FILING DATE: 1999-12-16
NUMBER OF SEQ ID NOS: 10451
SOFTWARE: pt_CT_genes Version 1.0
SEQ ID NO 1734
LENGTH: 4474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (89)...(3903)
OTHER INFORMATION: similar to gi559703 in the genepept database release 114,
OTHER INFORMATION: Run with FASTXY 3.3t00, default parameters
US-09-471-275-1734

Query Match 64.0%; Score 571.8; DB 56; Length 4474;
Best Local Similarity 82.2%; Pred. No. 2.5e-123;
Matches 657; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 4 atgatacagcaagtcacccctggaagatgccctgtccaaactgagacactgttgaagagctt 63
DB 143 atggcgccaggtgactctgagagcgctgtccaaactgagacactctctggagagctg 202
QY 64 cccctcccgacagcagcagctcatgagcctccaccttctccatcatgtgtaccaggtt 123
DB 203 cccctgcccagcagcagcctgcatgcagcccgccatctctgctctaccagcca 262
QY 124 aacttgacacaaacttgagacaggaatgcatttgtcacgggcatgtcaaggtacatt 183
DB 263 aatttcaacactaaacttgagacagaaatgcatttgttactggcatgcgaagatacatt 322
QY 184 gaggcagctacagtcaccctccagcatgaatgagatgctgagagaagacatgagtgcg 243
DB 323 gaacaagccaccgtccactctagatgaacagatgctggaggaggccaagaatgct 382
QY 244 gtcatgtgtacacctggcgagctgttcccgggccatttcccgaggtgaaatgcaacgag 303
DB 383 gtcatgtgtacacctggagagctgtctccggggccatcccccaggtgaaatgcaacgag 442

QY	304	cagcccaaccgagtagagatctatgagaagacagtagagtgctggagcggaggtcaacc	363
Db	443	cagcctaacagagtggaatactacgagaaaaccgtagggtctctggagcctgaggtcaca	502
QY	364	aagctcatgaagttcaatgtattttcagcgcaagccatcgagcgggttctcgacgcgaggctg	423
Db	503	aaactgatgaatttcattgtacttcagagaaatccattgagotgtctctcgggggaagtg	562
QY	424	aagcgctgtgccatccgacgagcgaggaagactttgtctctgaggcctaccctccctgacc	483
Db	563	aggcgctgtgccatgcccagagaggaaggaactctgtgcagaagcctaccctgatcaca	622
QY	484	cttggcaagtctcaatacattgttctgtctctgtagtagctaaagaacatgaagtgacgc	543
Db	623	ctgggcaaatctcaacatgtctcgtgtgtgcagcagctgaagaacatgaagtgcagt	682
QY	544	gtcaagaatgacacctcccctctacaagagggcagcacacatgttctctgcggaagatggcgagat	603
Db	683	gtgaagacgacacatcacgcgtacaaaggggccgcctcaagtttttctacgtaaaaatggcagat	742
QY	604	cccagctctatccagagtgctgcagaacctttccatgttctctggccaaccacaacaggatc	663
Db	743	ccacagtcctaccaggaatcgcagaatctgtccatgttctctggccaatcataacaagat	802
QY	664	accagtgctctcaaccagaacattgaagtgcctccagagctatgaggagctgctggctgac	723
Db	803	acacagctctctgcagcagcgcgcgaagtgaattcttggctacgaagagctcctggcagat	862
QY	724	attgtcaacatctgtgtggattactacgagagaacaagatgtacctgactccccagtgagaaa	783
Db	863	attgtgaatctgtgtgtggtattactacgagacaggatgtatttgcacgcccagtgagaaa	922
QY	784	catatgctctctcaaggtaa	802
Db	923	cacatgctctcaagtca	941

```

RESULT 13
US-60-172-360-19789
; Sequence 19789, Application US/60172360
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor
; FILE REFERENCE: GX-0007 P
; CURRENT APPLICATION NUMBER: US/60/172,360
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 29838
; SOFTWARE: PERL Program
; SEQ ID NO 19789
; LENGTH: 4723
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 375042.4
; US-60-172-360-19789

```

Query Match	64.0%	Score 571.8;	DB 92;	Length 4723;
Best Local Similarity	82.2%	-Pred. No. 2.6e-123;		
Matches	657: Conservative	0: Mismatches 142;	Indels	0: Gaps 0;

[illegible]

QY	124	aactttgacacaaacttttgagacagaaatgcatattgtcacaggccattgacagggtacatt	183
Db	215	aatctcaacactaaactttgaagacagaaatgcatattgttacaggccattgacagggtacatt	274
QY	184	gagcaggctacagtccactccacatgaatgagatgctggaggaagacatgagttatgcg	243
Db	275	gaacaagccaccgtccactctagatgaacgagatgctggaggggccaagaatatgct	334
QY	244	gtcatgctgtacacctggcgcaactttcccgggccattccccaggtgaaatgcaacgag	303
Db	335	gtcatgctgtacacctggagagagctgtcccgggccatccccagagtgaatgttaacgag	394
QY	304	cagcccaaccgagtagagatctatgagaagacagtagtgagtgctggagccgaggtcacc	363
Db	395	cagcctaacagagtgaatcactacgagaaaaccgtggaggtttctggagcctgaggtcaca	454
QY	364	aagctcatgaagttcatgtatttttcagcgcaagccatcgagcgtttctgcagcgagggtg	423
Db	455	aaactgatgaattctatgtacttccagagaaatgccattgagcgtttctgcgggggaagtg	514
QY	424	aagcggctgtgccatgccgagcgaggaagaaactttgtctctgagcctaccctcctgaac	483
Db	515	aggcgctgtgcccattgcccagagaggaagaaactctgttcgaagaagcctaccctgacaca	574
QY	484	cttggcaagttcatcaacatgtttgtctgctgagtgagctagaataaagaacatgaagtcagc	543
Db	575	ctgggcaaatctcaacatgtttcgtgtgtgacgagctgaagaacatgaagtcgagtc	634
QY	544	gtcaagaatgaccactccgctcacaagagggcagcagattcctgcgggaagatggcagat	603
Db	635	gtgaagaacgaccactcagctacaagaaggccgctcagtttttacctaaaaatggcagat	694
QY	604	cccagttatctccaggagctgcagaacctttcccatgttctctggcccaaccaacagagatc	663
Db	695	ccacagttccatccaggaaatcgagaaatctgccatgttccctggccaatcatacaagatc	754
QY	664	accagttgtctccaccagcaacttgaagtatccaggctatgaggagctcgtgctgac	723
Db	755	acacagttctctcagcagcagctcgaagtatttctgtctacgaagagctcctggcagat	814
QY	724	attgtcaacatctgtgtgattactacagaaacagatgtacctgactccccagtgagaaa	783
Db	815	attgtgaatctgtgtgtgattactacagaaacagatgtattgacgcccgagtgagaaa	874
QY	784	catatgctctccaagtaa	802
Db	875	cacatgcttctcaagtaa	893
RESULT 14			
US-09-270-767-14852			
; Sequence 14852, Application US/09270767			
; GENERAL INFORMATION:			
; APPLICANT: Homburger et al.			
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster			
; FILE REFERENCE: File Reference: 7326-094			
; CURRENT APPLICATION NUMBER: US/09/270,767			
; CURRENT FILING DATE: 1999-03-17			
; NUMBER OF SEQ ID NOS: 62517			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 14852			
; LENGTH: 1827			
; TYPE: DNA			
; ORGANISM: Drosophila melanogaster			
ns-09-270-767-14852			

Query Match 51.1%; Score 456; DB 45; Length 1827;
Best Local Similarity 73.1%; Pred. No. 2.1e-96;
Matches 585; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

Db	105	catgacggagaagattacgctagccgacgcgcgtgtccaaogtggaggtgttggacgagct	164
Qy	63	tcccctcccagacagcagccatgcacgcagctccacccttccccaatgtaccaggc	122
Db	165	atccctccgacgagcagccctgcacgcagcgacgcctgctcgattatctacaagc	224
Qy	123	taaccttgacacaaactttgagcacaggaatgcatttgcacgggcattgcaaggtacat	182
Db	225	aaacttcgatcacgaatttcgagatcgcaatggatttgcacgggtatcgccaagtacat	284
Qy	183	tgaacaggtcacagtcacatccacgcacatgaatgagatgctggaggaagaacatgagtatgc	242
Db	285	cgaaggagccaccaccatgcacaaactgaaatgtgctcctggacgagggcgagaacacgc	344
Qy	243	ggtaactctatacactggcgcagctgttcccgggcccattcccaggtgaaatacaacga	302
Db	345	agtaatgctctacacctggcgtgctgctcgcgcgcattcccgcagcccaagtcoccatga	404
Qy	303	gcagcccaacgcgagttagatctatgagaagacagtagaggtgctgcgcgcggaggtcac	362
Db	405	gcagccgaatcgcgagatctcaggaagacggtcgaggtgctagccccggaggcgaa	464
Qy	363	caagctcatgaattcatgtatttttcagcgaagcccatcgagcgttctgcagcgaggt	422
Db	465	caagctgctcaactcatctacttccaaacgcaaggcccatcgagggcttctccggcgaggt	524
Qy	423	gaagcggctgtgccatgccgcgacgcaggaaggactttgtctctgagggcctacctcctgcac	482
Db	525	gaagcgtgtgcccacgcgcgaagcgcgaagacttcgtgccgagcgctacctatgcac	584
Qy	483	cctggcaagtcatcaacatgtttgtcgtcgtctgtagagctaaagaacatgaagtgacag	542
Db	585	gctaggcaagtttatcaacatgtttgcgtgctggaagcgtgagaacatgaagtcacag	644
Qy	543	cgtacaagaatgaccactccgcctacaagaggcagacagcttccctcggaagatggcaga	602
Db	645	cgtgaagaacgattactcccactacagcgagctgcgcagttctctcaagtgtatgcgga	704
Qy	603	tcccagctctatccagagtgccagaaccttccatgttctctggcccaaccaacacagat	662
Db	705	ctcgacacacttgaggagtcgcgaatactgttccatgttctctggccacgcagaacagat	764
Qy	663	caccaggtgtctccaccagcaacttgaagtgtatcccaggctcatgaggagctgctggctga	722
Db	765	tcgcgacacgctcaaggacacgctggagaagatcgtgggctacgaggatctgctctcaga	824
Qy	723	catgtcaacatctgtgtggattactacgagacaagaatgtacctgactccagtcagaa	782
Db	825	tgtggtcaacatttgcgtgcatagttttcgagacaaagaatgtacttggccccctgaggaga	884
Qy	783	acatatgctcctcaaggtaa	802
Db	885	gcacatgctggtgtaaaagtta	904

RESULT 15

US-60-167-217-10144

US 00 107 217 10144
; Sequence 10144, Application US/60167217

; GENERAL INFORMATION:

APPLICANT: Li, Peter W. D.

TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC

1. TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL000152

; CURRENT APPLICATION NUMBER: US/

; CURRENT FILING DATE: 1999-1

; NUMBER OF SEQ ID NOS: 23195

```

; SOFTWARE: Fast

```

```

; SEQ ID NO 1014

```

; LENGTH: 39
TYPE: DNA

TYPE: DNA
ORGANISM: *Procorbilia*

; ORGANISM: Drosophila melanogaster

Query Match	51.18;	Score 456;	DB 90;	Length 3986;
-------------	--------	------------	--------	--------------

Best Local Similarity 73.18; Pred. No. 2.6e-96;

Matches 585; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

Qy	3	catgatcacgcacgctcaccttgaagatgacctgtccaaagtggacctgctgtgaagagct	67
Db	110	catgacggaagaattacgtatccgcgcgtgtccaaagtgaagtggtggacgagct	169
Qy	63	tccctcccgaccagcagcattgcatcgagctccaccttccctccatcatgtaccacggc	122
Db	170	atccctccgagacagcagccctgcatcgaggcgagccctgctcgattatctacaaggc	229
Qy	123	taacttgacacaaacttgaggacaggaatgcatttgtcacgggcatgtgcaaggtacat	182
Db	230	aaacttcgatcacgaatttcgaggatcgcaatggatttgtcacgggtatcgccaagtacat	289
Qy	183	tgacaggctacagtccactccagcatgaatgagatgctggagagaaggacatgagtatgc	242
Db	290	cgaggagccaccaccatgccaaacttgaatttgcgtctggacgaggggcagaaaaacgc	349
Qy	243	ggtcatcgttacacctggcgagctgttccgggccattccccaggtgaaatgcacaacga	302
Db	350	agtaatgctctacacttggcgtgctgctcgcgcccatctccgacgcccgaagtccaatga	409
Qy	303	gcagcccaaccagatgagatctatgagaagacagttagagtgctggagccggaggtccac	362
Db	410	gcagccgaatcgttgagatctacgagaagacggtcgaggtgctagccccggaggtgaa	469
Qy	363	caagctatgaagtcatgtatttttcagcgaaggccatcgagcgtgtctgcagcgaggt	422
Db	470	caagctgctcaactctatgtacttcccaacgaaaggccatcgagccctctcccggcaggt	529
Qy	423	gaagcggctgtgccatgccgagcgcaggaaggaccttgtctctgagggctacctctcgac	482
Db	530	gaagcgttgtgcacgcgagaagcgaaggacttcgtctcgaggcgtacctattgac	589
Qy	483	cttggcaagtctcaacatgtttgtgtctctggatgagctaaaagaacaatgaagtgcag	542
Db	590	gtcaggcaagtttatcaacatgtttgcgtgtcgtggacgagctgaagaacaatgaagtcag	649
Qy	543	cgtcaagaatgaccactccgcctacaagaggcgacacagttcctgcggaagatggcaga	602
Db	650	cgtgaagaacgattactccactacagcgagctgcygagttctctcaagtgatgctgga	709
Qy	603	tcccagttctacaggagtcgcgaagacccttccatgttctctggccaaacacacaggat	662
Db	710	ctcgacacacttgagagtcgcgaatactgtccatgttctctggccacgcagacaagat	769
Qy	663	caccagtgctccaccagaaacttgaagtgtatccccaggctatgaggagctgtctgctga	722
Db	770	tcgcgcacggtcgaagacacgctggagaagatcgttggtctacgaggtatgctctcaga	829
Qy	723	catgtcaacatctgttgattactacgagaaacaagatgtacctgactctccagtgagaa	782
Db	830	tgtggtcaacatttgcgtgcatatgttcgagacaaagatgtacttgacccctgaggagaa	889
Qy	783	acatatgctctcaaggtaa	802
Db	890	gcacatgctggtgataaagtta	909

Search completed: May 24, 2000, 01:01:42

search completed: M
Job time: 14274 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:23:02 ; Search time 6115.43 seconds
(without alignments)
-71.582 Million cell updates/sec

Title: US-09-215-435-116
Perfect score: 450
Sequence: 1 ctgctcagcgctgacgccg.....acggaacaagtctaggacag 450

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sts:*
- 14: gb_sy:*
- 15: gb_un:*
- 16: gb_vi:*
- 17: em_fun:*
- 18: em_humi:*
- 19: em_hum2:*
- 20: em_in:*
- 21: em_om:*
- 22: em_ov:*
- 23: em_ov:*
- 24: em_pat:*
- 25: em_ph:*
- 26: em_pl:*
- 27: em_ro:*
- 28: em_sts:*
- 29: em_sy:*
- 30: em_un:*
- 31: em_vi:*
- 32: gb_htg1:*
- 33: gb_htg2:*
- 34: gb_in1:*
- 35: gb_in2:*
- 36: em_ba1:*
- 37: em_ba2:*
- 38: em_hum3:*
- 39: em_hum4:*
- 40: gb_pr4:*
- 41: gb_htg3:*
- 42: gb_htg4:*
- 43: gb_htg5:*
- 44: gb_htg6:*

- 45: gb_htg7:*
- 46: em_htg1:*
- 47: em_htg2:*
- 48: em_htg3:*
- 49: em_hum5:*
- 50: gb_pl3:*
- 51: gb_pr5:*
- 52: gb_htg8:*
- 53: gb_htg9:*
- 54: gb_htg10:*
- 55: gb_htg11:*
- 56: gb_htg12:*
- 57: gb_htg13:*
- 58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	450	100.0	559	40	AF014955	40	AF014955 Homo sapi
2	324.8	72.2	381	12	AF161074	12	AF161074 Mus muscu
3	320.6	71.2	188683	53	AC010882	53	AC010882 Homo sapi
4	96.4	21.4	194723	54	AC008474	54	AC008474 Homo sapi
c	5	73.4	16.3	7218	5	I66494	I66494 Sequence 14
6	53.6	11.9	43346	54	AC011556	54	AC011556 Homo sapi
7	49.8	11.1	1009	13	CNS01JZO	13	AL147701 Anopheles
8	48.6	10.8	18765	2	AE000961	2	AE000961 Archaeogl
9	48.6	10.8	167439	44	AC020737	44	AC020737 Homo sapi
c	10	48	10.7	36281	55	AC023421	AC023421 Homo sapi
11	47.8	10.6	127520	56	AC023598	56	AC023598 Homo sapi
12	47.6	10.6	120379	44	AC011429	44	AC011429 Homo sapi
13	47.4	10.5	170820	57	AC016256	57	AC016256 Homo sapi
c	14	46.8	10.4	908	13	CNS01LOH	AL149026 Anopheles
15	46.6	10.4	63832	55	AC023655	55	AC023655 Homo sapi
16	46.4	10.3	947	13	CNS01LS7	13	AL146136 Anopheles
c	17	46.4	10.3	40481	43	AC018585	AC018585 Homo sapi
18	46.4	10.3	89346	43	AC017013	43	AC017013 Homo sapi
c	19	46.4	10.3	212071	58	AC020637	AC020637 Homo sapi
20	46.2	10.3	224448	32	PFMAL4P4	32	AL035477 Plasmodiu
21	46	10.2	59815	44	AC019096	44	AC019096 Homo sapi
22	45.6	10.1	45678	11	HSV351F8	11	Z70719 Human DNA s
c	23	45.6	10.1	73455	45	AC021589	AC021589 Homo sapi
24	45.6	10.1	169308	44	AC019101	44	AC019101 Homo sapi
c	25	45.6	10.1	227070	56	AC023888	AC023888 Homo sapi
26	45.4	10.1	70006	54	AC022671	54	AC022671 Homo sapi
c	27	45.4	10.1	197303	45	AC015583	AC015583 Mus muscu
28	45.2	10.0	78578	32	PFMAL13P9	32	AL096783 Plasmodiu
29	45.2	10.0	83709	54	AC022960	54	AC022960 Homo sapi
c	30	45.2	10.0	140350	43	AC015830	AC015830 Homo sapi
31	45.2	10.0	264399	42	AC015890	42	AC015890 Mus muscu
c	32	45	10.0	696	8	CNS01BD5	AL114081 Botrytis
33	45	10.0	895	13	CNS01KDS	13	AL148209 Anopheles
34	45	10.0	36471	9	D87021	9	D87021 Homo sapien
c	35	45	10.0	73020	54	AC022851	AC022851 Homo sapi
36	44.8	10.0	879	13	CNS01JRG	13	AL147405 Anopheles
37	44.8	10.0	5001	34	PFAABRA	34	J03902 plasmodium
38	44.8	10.0	115619	45	AC016068	45	AC016068 Homo sapi
c	39	44.8	10.0	121483	55	AC016238	AC016238 Homo sapi
40	44.8	10.0	175175	52	AC021538	52	AC021538 Homo sapi
c	41	44.8	10.0	190538	33	AL135926	AL135926 Homo sapi
42	44.6	9.9	43442	54	AC022426	54	AC022426 Homo sapi
c	43	44.6	9.9	142980	55	AC016304	AC016304 Homo sapi
44	44.6	9.9	222336	44	AC005308	44	AC005308 Plasmodiu
45	44.6	9.9	255925	44	AC020710	44	AC020710 Homo sapi

ALIGNMENTS

RESULT 1

AF014955 559 bp mRNA PRI 01-FEB-1999

LOCUS Homo sapiens TFAR19 mRNA, complete cds.

AF014955

ACCESSION AF014955.1 GI:2772828

VERSION

KEYWORDS human.

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Liu,H., Wang,Y., Zhang,Y., Song,Q., Di,C., Chen,G., Tang,J. and Ma,D.

TITLE TFAR19, a novel apoptosis-related gene cloned from human leukemia cell line TF-1, could enhance apoptosis of some tumor cells induced by growth factor withdrawal

JOURNAL Biochem. Biophys. Res. Commun. 254 (1), 203-210 (1999)

MEDLINE 99121216

REFERENCE

AUTHORS Liu,H.T., Wang,Y.G., Zhang,Y.M., Song,Q.S., Di,C.H., Yuan,Y. and Ma,D.L.

TITLE Direct Submission

JOURNAL Submitted (18-JUN-1997) Immunology, Beijing Medical University, No. 38, Xueyaun Road, Beijing 100083, Republic of China

REFERENCE

AUTHORS Liu,H.T., Wang,Y.G., Zhang,Y.M., Song,Q.S., Di,C.H., Yuan,Y. and Ma,D.L.

TITLE Direct Submission

JOURNAL Submitted (09-DEC-1997) Immunology, Beijing Medical University, No. 38, Xueyaun Road, Beijing 100083, Republic of China

REMARK Sequence update by submitter

REFERENCE

AUTHORS Liu,H.T., Wang,Y.G., Zhang,Y.M., Song,Q.S., Di,C.H., Yuan,Y. and Ma,D.L.

TITLE Direct Submission

JOURNAL Submitted (15-JAN-1998) Immunology, Beijing Medical University, No. 38, Xueyaun Road, Beijing 100083, Republic of China

REMARK Sequence update by submitter

COMMENT On Jan 16, 1998 this sequence version replaced gi:2668601.

FEATURES

source

1..559

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_line="TF-1"

1..559

/gene="TFAR19"

25..402

/gene="TFAR19"

/function="protein may function in the process of apoptosis"

/codon_start=1

/product="TFAR19"

/protein_id="AADI1579.1"

/db_xref="GI:2407068"

/translation="MADEELALRRRLAFLQAKHGDGDAAOQAEAKHREAEMRNL AQIDLSARLARLNLALVPEKTKAVENLIQARYQLSEKYSQGLTEILKKVVSQ TEXTTTVTFNRRKVMDSDEDDDI"

BASE COUNT 230 a 97 c 126 g 106 t

ORIGIN

Query Match 100.0%; Score 450; DB 40; Length 559;

Best Local Similarity 100.08; Pred. No. 1.9e-92;

Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgctccagcgtcgacgcgagccatgcgcgacgagagcttgaggcctagagagacag 60

Db 1 CTGCTCCAGCGCTGACGCGGACCCATGCGCGGACGAGAGCTTGAGCGGCTGAGGACAG 60

Qy 61 aggcctggccgaactcaggccaaacacgggatcctctgtatcgcgcccaacaggagca 120

Wed May 24 09:19:41 2000

```
QY 270 actaagtgaaggtatcagaacaaggtttaataagaataccttaaaaaagtaagccaaca 329
|||
Db 159135 AC-----TGAGAGGTCTCAGACACAGGTTTAAAGAAATCCTTTAAAGAAAGTAAGCCACAA 159190

QY 330 aacagaaaagacaacacagtgtaaatcaacagagaagaagtaattggactctgatgaaga 389
|||
Db 159191 AACAGAAAAGACAAACACAGTGAATTAACAGAGAAAGTAATGGACTCTGATGAAGA 159250

QY 390 tgacgattattgaactacaagtgcacagactagaaacttaacggaacaagcttaggaca 449
|||
Db 159251 AGACGATTATTGAACACAGAGTCTCAGAGACTAGAACTTACAGAACAAATCTTAGGACA 159310

QY 450 g 450
Db 159311 G 159311

RESULT 4
AC008474
LOCUS Homo sapiens chromosome 19 clone CTC-379B2, WORKING DRAFT SEQUENCE,
DEFINITION 22 unordered pieces.
AC008474
AC008474.3 GI:6899640
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 194723)
DOE Joint Genome Institute
Sequencing of Human Chromosome 19
Unpublished
2 (bases 1 to 194723)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 4, 2000 this sequence version replaced gi:6759078.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 153347 bases at least Q40
Consensus quality: 174602 bases at least Q30
Consensus quality: 184286 bases at least Q20
Estimated insert size: 194723; sum-of-contigs estimation
Estimated insert size: 182580; agarose-fp estimation
Quality coverage: 4.49x in Q20 bases; agarose-fp estimation
Quality coverage: 4.21x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1107: contig of 1107 bp in length
* gap of unknown length
* 1108 2198: contig of 1091 bp in length
* gap of unknown length
* 2199 3328: contig of 1130 bp in length
* gap of unknown length
* 3329 4995: contig of 1667 bp in length
* gap of unknown length
* 4996 6147: contig of 1152 bp in length
* gap of unknown length
* 6148 7783: contig of 1636 bp in length
* gap of unknown length
* 7784 9231: contig of 1448 bp in length
```

```
* 9232 13426: contig of 4195 bp in length
* 13427 17292: contig of 3866 bp in length
* 17293 21610: contig of 4318 bp in length
* 21611 28461: contig of 6851 bp in length
* 28462 34548: contig of 6087 bp in length
* 34549 42960: contig of 8412 bp in length
* 42961 50812: contig of 7852 bp in length
* 50813 60180: contig of 9368 bp in length
* 60181 72902: contig of 12722 bp in length
* 72903 87399: contig of 14497 bp in length
* 87400 99683: contig of 12284 bp in length
* 99684 119646: contig of 19963 bp in length
* 119647 144545: contig of 24899 bp in length
* 144546 169393: contig of 24848 bp in length
* 169394 194723: contig of 25330 bp in length.
FEATURES
location/Qualifiers
1..194723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTC-379B2"
BASE COUNT 51097 a 44698 c 45275 g 53622 t 31 others
ORIGIN
Query Match 21.4%; Score 96.4; DB 54; Length 194723;
Best Local Similarity 99.0%; Pred. No. 3.7e-12;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 353 aattcaacagagaagaaagtaattggactctgatgaagatgacgattattgaactacaagt 412
|
Db 78806 AGTTCAACAGAGAAAGTAATGGACTCTGATGAAGATGACGATTATTGAACTACAAGTG 78865
|
QY 413 ctccagactagaacttaacggaacaagcttaggacag 450
|
Db 78866 CTCACAGACTAGNACTTACGGAACAAGTCTAGGACAG 78903
|
RESULT 5
166494/c 7218 bp DNA PAT 28-DEC-1997
LOCUS Sequence 14 from patent US 5670367.
DEFINITION
AC008474
AC008474.1 GI:2724471
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 7218)
AUTHORS
Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE
Recombinant fowlpox virus
JOURNAL
Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
```

```
Query Match      16.3%; Score 73.4; DB 5; Length 7218;
Best Local Similarity 6.1%; Pred. No. 5e-07;
Matches 26; Conservative 240; Mismatches 161; Indels 0; Gaps 0;

QY 23 ccatggcgacagaggcttgagcgctgagagagacagagctgcccagctgcagcgcca 82
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1435 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1376
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 83 aacacgggagctctgtgatcgccgcccacagagaaagacacagggagcagaatga 142
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1375 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 143 gaacacagtattagccacagttctgacagtcggccgcccaggttaagtaactag 202
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1315 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 203 cactgttaagcctgaaacactaaagcagtagagaaattacccttatcacagatgcca 262
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1255 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 263 atggacaactaagtgaaggtatcagaacaaggtttaatagaaatccttaaaaaagtaa 322
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1195 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 323 gccacaacagaaaagacaaacagtgtaaatcaacagagaagaagtaatggactcg 382
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1135 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 383 atgaagatgacgattattgaatacaagtcacagactagaacttaacggaacagtc 442
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1075 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1016
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 taggaca 449
| : : : : :
Db 1015 TATGGCA 1009

RESULT 6
AC011556 43346 bp DNA HTG 06-FEB-2000
LOCUS Homo sapiens chromosome 19 clone LLNLR-281B11, WORKING DRAFT
DEFINITION SEQUENCE, 6 ordered pieces.
AC011556
AC011556.3 GI:6910518
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
DOE Joint Genome Institute.
Sequencing of Human Chromosome 19
Unpublished
JOURNAL
TITLE
AUTHORS
REFERENCE 1 (bases 1 to 43346)
DOE Joint Genome Institute.
Direct Submission
Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 6, 2000 this sequence version replaced gi:6693128.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
```

```
* the accession number will be preserved.
* 1 3707: contig of 3707 bp in length
* 3708 12743: gap of unknown length
* 12744 30266: contig of 17523 bp in length
* 30267 32440: contig of 2174 bp in length
* 32441 39703: gap of unknown length
* 39704 43346: contig of 7263 bp in length
* Location/Qualifiers
  1. 43346
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="19"
  /clone="LLNLR-281B11"
BASE COUNT 9538 a 12146 c 12398 g 8958 t 306 others
ORIGIN

Query Match      11.9%; Score 53.6; DB 54; Length 43346;
Best Local Similarity 54.6%; Pred. No. 0.017;
Matches 107; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 208 gtaaacctgaaacactaaagcagtagagaattacccttatcacagatgccaagatgga 267
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 39943 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 40002
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 268 caactaagtgagaggtatcagacaaggtttatagaaatccttaaaaaagtaagccaa 327
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 40003 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 40062
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 caaacagaaagacacacacagtcgaattcaacagagaagaagtaatgactctgatgaa 387
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 40063 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 40122
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 gatgacgattattgaa 403
| : : : : :
Db 40123 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 40138

RESULT 7
CNS01JZO 1009 bp DNA STS 17-FEB-2000
LOCUS Anopheles gambiae STS T7 end of clone 14M14 of NotreDamel library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
sequence tagged site.
AC0147701
AC0147701.1 GI:7005847
KEYWORDS STS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Culicidae; Anopheles.
REFERENCE 1 (bases 1 to 1009)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
REFERENCE 2 (bases 1 to 1009)
AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
Location/Qualifiers
```



```
identity: 33.14; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="thymidylate synthase, putative"
/protein_id="AAB89199.1"
/db_xref="GI:2648479"
/transl_table=11
/product="MYANTDPIIIRARRKASSLADFFSICLAMLKRFSSMADKTFPHGK
LFNLRLKWLMLTTPPEAKEMLVSKVIGGETCSRFGDYRLSKPTMVVVEPNSFG
FFEDVDCCKEKSERLSRCSVESAEKLRKSPHTRASIPNTPKDLHCLNPAATREIS
FIFHEKLHUTALSRMECLSYFEHNFDFLVALETICRKTGMEEGSIGMLIAVPHFE
ROVERALSYSGLRETYGHELGTHLVEDYISSAWHSALETIYTNKKRTWMDIFE
GOKESYFVRLVELEKPEHLDKAPTEKYGEIYAHDIYIMHAKLDGEVRRSLIK
EGEETVADRARVCDRDDVKVDQYKVIKELKEDSCRRDCCYVVISRPWLLSDEPCL
RGYQSKYDGLLGTFTFMSNDAYGAMHANMYAFALLTKYVAELGTGFKSYRYNHFALD
AHYAEFFDAVREILYPESPILDVSGKG"
3438..4361
/gene="AF2048"
CDS
3438..4361
/gene="AF2048"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF2048"
/protein_id="AAB89212.1"
/db_xref="GI:2648492"
/transl_table=11
/product="MAKQMEKKSRLLAFLLALIMIGSVFAYMLSGGSAHEHREVVRL
DDREYVNTPADPVYQYVYNSYSLGSKDPPLASMTVTDLOKLIPAFSRQVLEV
TRGISQVMTDGETVPLFYVDAGMSKIYFAKEDIKHGNTLQVRRPGIALVSELS
LVVGYLPEKAVDVEGNSPSFNKTYSLSRINGSFAYAFAYGDVVKQWTRVGN
SPADPFEGYRYFNNSSEKVMAMHFEYFEGGNESEKNFEYKYVQNFQDGLSVA
VMEDKNFTKVVNARNPILTWQISFNNTQNES"
4351..5487
/gene="AF2049"
CDS
4351..5487
/gene="AF2049"
/notes="similar to GB:L77117 PID:L591954 percent identity:
31.36; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB89198.1"
/db_xref="GI:2648478"
/transl_table=11
/product="MRVEKGTALAKGRVRIKGEAVCGASISEFFTEKFEVPIYCLDD
CEILVEGGRVDGCTIPESWKKLAKRDWEVFLGCVGVGKSTLATYLANLGGCYI
LDLDIGQSELLHGAMGYFARNVNLQAEYINGYFVGVISPMKEVKCIRGVARLW
GELKLEGRKIVNTTGWVGARALDYKLAKLEVINPDIVASFEDANLEKDVNFRVBSG
EVVERERERAKIRMEKYRELEGLSKLVVDKSRVSGRFNGKNIVKDFIEDVLEVRV
KEVRKGEDEFLILTEKIEVDSSIIKSLKMLVEVEDVYVVAEMEELGVLVGVYRGSY
LGIGVYRFRGSEFFVETIETHDFDRIVLGEFGIAQGEYMLRLP"
5495..5875
/gene="AF2050"
CDS
5495..5875
/gene="AF2050"
/notes="similar to GB:L77117 PID:L591729 percent identity:
42.74; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB89204.1"
/db_xref="GI:2648491"
/transl_table=11
/product="MELTKTAKRVEIIDIOTVERCVRSDGLVLVYTPHTTALVI
NGERGLLEDILEFMEKLVPGYKGYKHDRLDSNADHLKATLLGNSVVVPVSGKLAL
GTWQRIFLFEGDPGTRRVIRVAL"
5872..6447
/gene="AF2051"
CDS
5872..6447
/gene="AF2051"
/notes="similar to GB:M83138 SP:P28692 GB:U01376 PID:L46027
PID:606117 percent identity: 40.76; identified by sequence
similarity; putative"
/codon_start=1
```

```
/transl_table=11
/product="cell division protein (ftsJ)"
/protein_id="AAB89203.1"
/db_xref="GI:2648490"
/transl_table=11
/product="MKDRQDIYYWKAKEGYRSRAAYKLLQNNRTFKLIREGLVLDL
GATPGSQVAALLGAVRVAVDINPMKPLENTVTFIRGDTLPLETLEKIREISPDYDV
MSDAPKISGKWTIDHLRSIDLARASFSTAKEVLKPGGNFVVKVFGQEEIQKPFNELK
PHFRPKFHSPOASRRKRSAEVYFIGRFRKI"
complement(6434..7270)
/gene="AF2052"
CDS
complement(6434..7270)
/gene="AF2052"
/notes="similar to PID:L197361 percent identity: 42.65;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="2-ketolisovalerate ferredoxin oxidoreductase,
subunit beta (vorB)"
/protein_id="AAB89197.1"
/db_xref="GI:2648477"
/transl_table=11
/product="MILPGNASCQGPISMMRCLDEIENPVLVIPAGCSTVIAGLHP
KSARKCPVHVPPASTHVSACGLSEAYKILGVGATVVMNGDGAADIGFAALSASATP
KODILTVVVDNEAYMNTGTQSSLTFWKARTTSPKGTEDRRNLALVMLAHKVDYVA"
Query Match 10.8%; Score 48.6; DB 2; Length 18765;
Best Local Similarity 50.5%; Pred. No. 0.22;
Matches 145; Conservative 0; Mismatches 139; Indels 3; Gaps 1;
QY 34 gagagagcttgagcgctgagagacagagctgcccagctgcagggccaaacacgggat 93
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16953 GATGATCTGAGGAGAAATAAGCGCAGGAGCTGATGAGTTGCAG---AGGCAGAAGGAG 16897
QY 94 cctggtgatcgcccccaacaggaagcaaacacaggaagcagaaatgagaaacagtatc 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16896 CTTGAGGAGCTGCAGAGGAGGAGATGAGGACAGCTTGAGGCACAGAGAGGCAATT 16837
QY 154 ttgcccgaagtcttgatcatcgccgcccggccaggttaagtaacttagcacttgtaag 213
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16836 TTGNGGCAATCTTGAGCCGAGCTAAGGAGAGACTTTCACGCTTAAAGCTCGCTCAT 16777
QY 214 cctgaaaaaataaagcagtagaataaccttatcacagatggcaagatggcaacta 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16776 CCTCAATTTGCTAGGCAGTTGAGAACACAGCTTATTTATCTGCTCAGGCTGGAAGAATT 16717
QY 274 agtgagaaggtatcacagaacaaagtttaatagaaatccttaaaaaagt 320
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16716 CAGAGCAAGATACTGATAAAATCTCGTTGAAATTTTGAAGAGAT 16670
RESULT 9
AC020737 167439 bp DNA HTG 08-JAN-2000
LOCUS Homo sapiens clone RP11-75009, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC020737
ACCESSION AC020737
VERSION AC020737.1 GI:6682696
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 167439)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
JOURNAL 2 (bases 1 to 167439)
Waterston,R.H.
AUTHORS Direct Submission
TITLE Submitted (08-JAN-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This record contains 235 individual
* sequencing reads that have not been assembled into
```

* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 668: contig of 668 bp in length
* 669
* 1184: contig of 506 bp in length
* 1185
* 1194: gap of unknown length
* 1195
* 1508: contig of 314 bp in length
* 1509
* 1518: gap of unknown length
* 1519
* 1792: contig of 274 bp in length
* 1793
* 1802: gap of unknown length
* 1803
* 2476: contig of 674 bp in length
* 2477
* 2486: gap of unknown length
* 2487
* 3155: contig of 669 bp in length
* 3156
* 3165: gap of unknown length
* 3166
* 3810: contig of 645 bp in length
* 3811
* 3820: gap of unknown length
* 3821
* 4484: contig of 664 bp in length
* 4485
* 5130: contig of 636 bp in length
* 5131
* 5140: gap of unknown length
* 5141
* 6200: contig of 1060 bp in length
* 6201
* 6210: gap of unknown length
* 6211
* 6878: contig of 668 bp in length
* 6879
* 6888: gap of unknown length
* 6889
* 7277: contig of 839 bp in length
* 7278
* 7737: gap of unknown length
* 7738
* 8398: contig of 661 bp in length
* 8399
* 8408: gap of unknown length
* 8409
* 8850: contig of 442 bp in length
* 8851
* 8960: gap of unknown length
* 8961
* 9627: contig of 767 bp in length
* 9628
* 9637: gap of unknown length
* 9638
* 10282: contig of 645 bp in length
* 10283
* 10292: gap of unknown length
* 10293
* 10957: contig of 665 bp in length
* 10958
* 10967: gap of unknown length
* 10968
* 11635: contig of 668 bp in length
* 11636
* 11645: gap of unknown length
* 11646
* 12569: contig of 924 bp in length
* 12570
* 12579: gap of unknown length
* 12580
* 13797: contig of 1218 bp in length
* 13798
* 13807: gap of unknown length
* 13808
* 14471: contig of 664 bp in length
* 14472
* 14481: gap of unknown length
* 14482
* 15160: contig of 679 bp in length
* 15161
* 15170: gap of unknown length
* 15171
* 15403: contig of 233 bp in length
* 15404
* 15413: gap of unknown length
* 15414
* 16132: contig of 719 bp in length
* 16133
* 16142: gap of unknown length
* 16143
* 16801: contig of 659 bp in length
* 16802
* 16811: gap of unknown length
* 16812
* 17109: contig of 298 bp in length
* 17110
* 17119: gap of unknown length
* 17120
* 17788: contig of 669 bp in length
* 17789
* 17798: gap of unknown length
* 17799
* 18468: contig of 670 bp in length
* 18478: gap of unknown length
* 18479
* 19140: contig of 668 bp in length
* 19141
* 19156: gap of unknown length
* 19157
* 19825: contig of 669 bp in length
* 19826
* 19835: gap of unknown length
* 19836
* 20514: contig of 679 bp in length
* 20515
* 20524: gap of unknown length
* 20525
* 21188: contig of 664 bp in length
* 21189: gap of unknown length

21199
* 21851
* 21860: gap of unknown length
* 21861
* 22708: contig of 848 bp in length
* 22709
* 22718: gap of unknown length
* 22719
* 23483: contig of 765 bp in length
* 23484
* 23493: gap of unknown length
* 23494
* 24166: contig of 673 bp in length
* 24167
* 24176: gap of unknown length
* 24177
* 24818: contig of 642 bp in length
* 24819
* 24828: gap of unknown length
* 24829
* 25863: contig of 1035 bp in length
* 25864
* 25873: gap of unknown length
* 25874
* 26559: contig of 686 bp in length
* 26560
* 26569: gap of unknown length
* 26570
* 27634: contig of 1065 bp in length
* 27635
* 27644: gap of unknown length
* 27645
* 27964: contig of 320 bp in length
* 27965
* 27974: gap of unknown length
* 27975
* 28656: contig of 682 bp in length
* 28657
* 28666: gap of unknown length
* 28667
* 29425: contig of 759 bp in length
* 29426
* 29436: gap of unknown length
* 29437
* 30566: contig of 1131 bp in length
* 30567
* 30576: gap of unknown length
* 30577
* 31096: contig of 520 bp in length
* 31097
* 31768: contig of 662 bp in length
* 31769
* 31778: gap of unknown length
* 31779
* 32458: contig of 680 bp in length
* 32459
* 32468: gap of unknown length
* 32469
* 32949: contig of 481 bp in length
* 32950
* 32959: gap of unknown length
* 32960
* 33676: contig of 717 bp in length
* 33677
* 33686: gap of unknown length
* 33687
* 34351: contig of 665 bp in length
* 34352
* 34361: gap of unknown length
* 34362
* 34858: contig of 497 bp in length
* 34859
* 34868: gap of unknown length
* 34869
* 35440: contig of 672 bp in length
* 35441
* 35550: gap of unknown length
* 35551
* 35863: contig of 313 bp in length
* 35864
* 35873: gap of unknown length
* 35874
* 36338: contig of 465 bp in length
* 36339
* 36348: gap of unknown length
* 36349
* 37003: contig of 655 bp in length
* 37004
* 37013: gap of unknown length
* 37014
* 37684: contig of 671 bp in length
* 37685
* 37694: gap of unknown length
* 37695
* 38384: contig of 690 bp in length
* 38385
* 38394: gap of unknown length
* 38395
* 38800: contig of 406 bp in length
* 38801
* 38810: gap of unknown length
* 38811
* 39489: contig of 679 bp in length
* 39490
* 39499: gap of unknown length
* 39500
* 39793: contig of 294 bp in length
* 39794
* 39803: gap of unknown length
* 39804
* 40097: contig of 294 bp in length
* 40098
* 40107: gap of unknown length
* 40108
* 40760: contig of 653 bp in length
* 40761
* 40770: gap of unknown length
* 40771
* 41436: contig of 666 bp in length
* 41437
* 41446: gap of unknown length
* 41447
* 42115: contig of 669 bp in length
* 42116
* 42125: gap of unknown length
* 42126
* 42381: contig of 256 bp in length
* 42382
* 42391: gap of unknown length
* 42392
* 43068: contig of 677 bp in length
* 43069
* 43078: gap of unknown length
* 43079
* 43739: contig of 661 bp in length
* 43740
* 43749: gap of unknown length
* 43750
* 44403: contig of 654 bp in length
* 44404
* 44413: gap of unknown length
* 44414
* 44890: contig of 477 bp in length

10014 10740: contig of 727 bp in length
gap of unknown length
10741 11446: contig of 706 bp in length
gap of unknown length
11447 12180: contig of 734 bp in length
gap of unknown length
12181 12904: contig of 724 bp in length
gap of unknown length
12905 13616: contig of 712 bp in length
gap of unknown length
13617 14345: contig of 729 bp in length
gap of unknown length
14346 15044: contig of 699 bp in length
gap of unknown length
15045 15732: contig of 688 bp in length
gap of unknown length
15733 16458: contig of 726 bp in length
gap of unknown length
16459 17185: contig of 727 bp in length
gap of unknown length
17186 17879: contig of 694 bp in length
gap of unknown length
17880 18582: contig of 703 bp in length
gap of unknown length
18583 19281: contig of 699 bp in length
gap of unknown length
19282 20000: contig of 719 bp in length
gap of unknown length
20001 20690: contig of 690 bp in length
gap of unknown length
20691 21396: contig of 706 bp in length
gap of unknown length
21397 22100: contig of 704 bp in length
gap of unknown length
22101 22811: contig of 711 bp in length
gap of unknown length
22812 23534: contig of 723 bp in length
gap of unknown length
23535 24269: contig of 735 bp in length
gap of unknown length
24270 24984: contig of 715 bp in length
gap of unknown length
24985 25677: contig of 693 bp in length
gap of unknown length
25678 26372: contig of 695 bp in length
gap of unknown length
26373 27075: contig of 703 bp in length
gap of unknown length
27076 27788: contig of 713 bp in length
gap of unknown length
27789 28521: contig of 733 bp in length
gap of unknown length
28522 29213: contig of 692 bp in length
gap of unknown length
29214 29941: contig of 728 bp in length
gap of unknown length
29942 30651: contig of 710 bp in length
gap of unknown length
30652 31352: contig of 701 bp in length
gap of unknown length
31353 32062: contig of 710 bp in length
gap of unknown length
32063 32773: contig of 711 bp in length
gap of unknown length
32774 33466: contig of 693 bp in length
gap of unknown length
33467 34181: contig of 715 bp in length
gap of unknown length
34182 34898: contig of 717 bp in length
gap of unknown length
34899 35567: contig of 669 bp in length
gap of unknown length
35568 36281: contig of 714 bp in length

FEATURES
source
Location/Qualifiers
1..36281
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-116018"
/clone.lib="RPC1-11 Human Male BAC"
BASE COUNT 9003 a 7259 c 6967 g 12196 t 856 others
ORIGIN
Query Match 10.7%; Score 48; DB 55; Length 36281;
Best Local Similarity 44.4%; Pred. NO. 0.31;
Matches 159; Conservative 0; Mismatches 199; Indels 0; Gaps 0;
QY 16 cgcgcagccatggcgacagagagcttgagcgctgagcgagagagagagcgctgagcgagctg 75
|||||
Db 2304 CGCGGAATTCTGAGGAGCTTGCATAATTTCTGTGCTGACGCGAGAGTCTCCTTAGGACC 2245
QY 76 cagcccaaacacgggatcctgtgatcgcccaacaggaagcagagcagcgagagagcga 135
|||||
Db 2244 CCAGCAGAACTGGGCTTCTTGAGTTAAGCTCTTCTCAGAAAAATAAGGAAAGGCGTCA 2185
QY 136 gaaatgagaaacagatctcttagcccaagttctgatcgatcgcccgcccgaggttaagt 195
Db 2184 NCTCAGCAGNNATNCGGNGNCTCCGAANNNTAGTCTTCCCGCCCGCCCGTGGGAAAT 2125
QY 196 aacttagcacttgaaagcctgaaagcctgaaagcctgaaagcctgaaagcctgaaagcct 255
Db 2124 AAAT 2065
QY 256 gcaagatatggacaaactaagtggagaggtatcgacacagtgaaattcaacagagagaa 315
Db 2064 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 2005
QY 316 aaagtaagcccaacaaacag 373
Db 2004 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 1947
RESULT 11
AC023598/c 127520 bp DNA HTG 19-FEB-2000
LOCUS Homo sapiens chromosome 3 clone RP11-221E20, *** SEQUENCING IN
DEFINITION PROGRESS ***, 77 unordered pieces.
AC023598
AC023598.2 GI:6996882
VERSION HTG: HTGS_PHASE1.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 127520)
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R.,
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wambach,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission

Unpublished
2 (bases 1 to 127520)

Worley,K.C.

Direct Submission

Submitted (16-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Feb 18, 2000 this sequence version replaced gi:6980130.

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HAGQ

Center clone name: RP11-221E20

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.980611

Consensus quality: 67656 bases at least Q40

Consensus quality: 88015 bases at least Q30

Consensus quality: 98162 bases at least Q20

Estimated insert size: 109398; sum-of-contigs estimation

Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 77 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1
1183 1182: contig of 1182 bp in length
1203 1202: gap of unknown length
1206 1206: contig of 784 bp in length
1207 2006: gap of unknown length
2007 2804: contig of 798 bp in length
2805 2824: gap of unknown length
2825 3006: contig of 782 bp in length
3007 3626: gap of unknown length
3627 4366: contig of 740 bp in length
4367 4386: gap of unknown length
4387 5041: contig of 655 bp in length
5042 5061: gap of unknown length
5062 5817: contig of 756 bp in length
5818 5837: gap of unknown length
5838 6577: contig of 740 bp in length
6578 6597: gap of unknown length
6598 7608: contig of 1011 bp in length
7609 7628: gap of unknown length
7629 8395: contig of 767 bp in length
8396 8415: gap of unknown length
8416 9197: contig of 782 bp in length
9198 9217: gap of unknown length
9218 9382: contig of 765 bp in length
9383 10002: gap of unknown length
10003 11066: contig of 1064 bp in length
11067 11086: gap of unknown length
11087 11827: contig of 741 bp in length
11828 11847: gap of unknown length
11848 12894: contig of 1047 bp in length
12895 12914: gap of unknown length
12915 14057: contig of 1143 bp in length
14058 14077: gap of unknown length
14078 14801: contig of 724 bp in length
14802 14821: gap of unknown length
14822 15844: contig of 1023 bp in length
15845 15864: gap of unknown length
15865 16870: contig of 1006 bp in length
16871 16890: gap of unknown length

16891 18018: contig of 1128 bp in length
18019 18038: gap of unknown length
18039 18925: contig of 887 bp in length
18926 18945: gap of unknown length
18946 19933: contig of 988 bp in length
19934 19953: gap of unknown length
19954 20729: contig of 776 bp in length
20730 20749: gap of unknown length
20750 21961: contig of 1212 bp in length
21962 21981: gap of unknown length
21982 23101: contig of 1120 bp in length
23102 23121: gap of unknown length
23122 24215: contig of 1094 bp in length
24216 24235: gap of unknown length
24236 25159: contig of 924 bp in length
25160 25179: gap of unknown length
25180 25945: contig of 766 bp in length
25946 26703: contig of 738 bp in length
26704 26723: gap of unknown length
26724 27490: contig of 767 bp in length
27491 27510: gap of unknown length
27511 28776: contig of 1266 bp in length
28777 28796: gap of unknown length
28797 30040: contig of 1244 bp in length
30041 30060: gap of unknown length
30061 31243: contig of 1183 bp in length
31244 31263: gap of unknown length
31264 32578: gap of unknown length
32579 33366: contig of 788 bp in length
33367 33386: gap of unknown length
33387 34463: contig of 1077 bp in length
34464 34483: gap of unknown length
34484 35692: contig of 1209 bp in length
35693 35712: gap of unknown length
35713 36936: contig of 1224 bp in length
36937 36956: gap of unknown length
36957 38115: contig of 1159 bp in length
38116 38135: gap of unknown length
38136 39556: contig of 1421 bp in length
39557 39577: gap of unknown length
39578 41094: contig of 1518 bp in length
41095 41114: gap of unknown length
41115 42196: contig of 1082 bp in length
42197 42216: gap of unknown length
42218 43530: contig of 1314 bp in length
43531 43550: gap of unknown length
43551 44428: contig of 878 bp in length
44429 44448: gap of unknown length
44449 46102: contig of 1634 bp in length
46103 46122: gap of unknown length
46123 47906: contig of 1784 bp in length
47907 47926: gap of unknown length
47927 49750: contig of 1824 bp in length
49751 49770: gap of unknown length
49771 51452: contig of 1682 bp in length
51453 51472: gap of unknown length
51473 53129: contig of 1657 bp in length
53130 53149: gap of unknown length
53150 54827: contig of 1678 bp in length
54828 54847: gap of unknown length
54848 57212: contig of 2365 bp in length
57213 57232: gap of unknown length
57233 58840: contig of 1608 bp in length
58841 58860: gap of unknown length
58861 60161: contig of 1301 bp in length
60162 60181: gap of unknown length
60182 62146: contig of 1965 bp in length
62147 63974: contig of 1808 bp in length
63975 63994: gap of unknown length
63995 65455: contig of 1461 bp in length

* 65456		65475: gap of unknown length
* 65476		67429: contig of 1954 bp in length
* 67430		67449: gap of unknown length
* 67450		68488: contig of 1039 bp in length
* 68489		68508: gap of unknown length
* 68509		71445: contig of 2937 bp in length
* 71446		71465: gap of unknown length
* 71466		72788: contig of 1323 bp in length
* 72789		72808: gap of unknown length
* 72809		74949: contig of 2141 bp in length
* 74950		74969: gap of unknown length
* 74970		77956: contig of 2987 bp in length
* 77957		77976: gap of unknown length
* 77977		80307: contig of 2331 bp in length
* 80308		80327: gap of unknown length
* 80328		83135: contig of 2808 bp in length
* 83136		83155: gap of unknown length
* 83156		85464: contig of 2309 bp in length
* 83465		85484: gap of unknown length
* 85485		87235: contig of 1751 bp in length
* 87236		87255: gap of unknown length
* 87256		89925: contig of 2670 bp in length
* 89926		89945: gap of unknown length
* 89946		92426: contig of 2481 bp in length
Query Match		10.6%; Score 47.8; DB 56; Length 127520;
Best Local Similarity		39.8%; Pred. No. 0.37;
Matches 157; Conservative		0; Mismatches 237; Indels 0; Gaps 0;
Qy	28	gcgcagcaggagcttgagcgctgagagacagagagctgcccagctgcaggccaacac 87
Db	960	GTGGAGACCCAGCAAGTACGTTAACGCGACAGCGCACGCGGAGCGGAGAGAGGC 901
Qy	88	ggggatcctggtgatcgggcccaacaggaaagcaagacacagaggaagcagagaatggaac 147
Db	900	GAGAGAAGCGGAG 841
Qy	148	agtatctagcccaagttctgagtcagtcagtcgcccggcggcaggttaagtaacttagcactt 207
Db	840	NAAAG 781
Qy	208	gtaagcctgaaacaaactaaagcagtagaggaattacctatcacagatggcagatgga 267
Db	780	AAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 721
Qy	268	caactaagtgaaggttatcagaacaaggtttaaataagaaatccttaaaaaagtaagccaa 327
Db	720	AAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 661
Qy	328	caacagaaagacaaacaaacagtgaaattccacagagaaagtaagtgaactctgatgaa 387
Db	660	AAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 601
Qy	388	gatgacgattattgaactacaagtgtccacagac 421
Db	600	CCATATATATACATATATACATATACATATACATATACATATACATATACATAT 567
RESULT 12		
AC011429		
LOCUS		AC011429 120379 bp DNA HTG 19-DEC-1999
DEFINITION		Homo sapiens chromosome 5 clone P1_545G1, WORKING DRAFT SEQUENCE, 25 unordered pieces.
ACCESSION		AC011429
VERSION		AC011429.2 GI:6604399
KEYWORDS		HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE		human.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS		1 (bases 1 to 120379)
TITLE		DOE Joint Genome Institute. Sequencing of Human Chromosome 5

JOURNAL REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 120379)
TITLE	DOE Joint Genome Institute.
JOURNAL	Direct Submission
COMMENT	Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA On Dec 20, 1999 this sequence version replaced gi:6013513. -----Genome Center Center: Joint Genome Institute Web site: http://www.jgi.doe.gov -----Summary Statistics Estimated insert size: 120379; sum-of-contigs estimation Estimated insert size: 80000; pulse field gel estimation ----- * NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 1017: contig of 1017 bp in length * 1018 2125: gap of unknown length * 2126 3153: contig of 1108 bp in length * 3154 4187: gap of unknown length * 4188 5191: contig of 1028 bp in length * 5192 6343: gap of unknown length * 6344 7558: contig of 1034 bp in length * 7559 9384: gap of unknown length * 9385 10876: contig of 1004 bp in length * 10877 12399: gap of unknown length * 12400 14703: contig of 1152 bp in length * 14704 16578: gap of unknown length * 16579 17929: contig of 1215 bp in length * 17930 19883: gap of unknown length * 19884 21385: contig of 1826 bp in length * 21386 23222: gap of unknown length * 23223 24703: contig of 1492 bp in length * 24704 26381: gap of unknown length * 26382 28141: contig of 1523 bp in length * 28142 29484: gap of unknown length * 29485 31854: contig of 1351 bp in length * 31855 34803: gap of unknown length * 34804 38074: contig of 1937 bp in length * 38075 75014: gap of unknown length * 75015 120379: contig of 1381 bp in length * Location/Qualifiers 1. 120379

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="p1_545G1"
BASE COUNT 33442 a 26638 c 26050 g 34159 t 90 others
ORIGIN

Query Match 10.6%; Score 47.6; DB 44; Length 120379;
Best Local Similarity 53.2%; Pred. No. 0.41;
Matches 101; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 210 aagcctgaaacaaacagcagtagagaattacccttatacagatgcaagatagcaca 269
Db 23737 AAAAATTAAAAAATAAATAATATAAATAAATTTTAAAAATAAAAAATAAAAA 23796

QY 270 actaagtggagaggtatcagacaaagggttaataagaaatccttaaaaaagtagcaca 329
Db 23797 AATATTATATATAAATGATATAAATAAATAAATAAATAAATAAATAAATAAAT 23856

QY 330 aacagaaagacacacagtgaaattcaacagacagaaagaaagtaagactctgataaga 389
Db 23857 AATAAATTATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 23916

QY 390 tgacgattat 399
Db 23917 ATATGAATAT 23926

RESULT 13
AC016256
LOCUS
DEFINITION Homo sapiens clone RP11-98E6, *** SEQUENCING IN PROGRESS ***, 27
unordered pieces.
ACCESSION AC016256
VERSION AC016256.2 GI:7007812
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 170820)
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Barbara,J., Blankenburg,K.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Guevara-Wang,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabnah,M.,
Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 170820)
Worley,K.C.
Direct Submission
Submitted (24-NOV-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 19, 2000 this sequence version replaced gi:6466907.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMOX
Center clone name: RP11-98E6
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 143268 bases at least Q40
Consensus quality: 153640 bases at least Q30
Consensus quality: 158626 bases at least Q20
Estimated insert size: 170820; agarose-fp estimation
Estimated insert size: 166163; sum-of-contigs estimation
Quality coverage: 2.9x in Q20 bases; agarose-fp estimation
Quality coverage: 2.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 820: contig of 820 bp in length
* 821 840: gap of unknown length
* 841 1645: contig of 805 bp in length
* 1646 1665: gap of unknown length
* 1666 2484: contig of 819 bp in length
* 2485 2504: gap of unknown length
* 2505 3282: contig of 778 bp in length
* 3283 3302: gap of unknown length
* 3303 4195: contig of 893 bp in length
* 4196 4215: gap of unknown length
* 4216 5537: contig of 1322 bp in length
* 5538 9321: contig of 3764 bp in length
* 9322 9341: gap of unknown length
* 9342 11925: contig of 2584 bp in length
* 11926 11945: gap of unknown length
* 11946 16630: contig of 4685 bp in length
* 16631 16650: gap of unknown length
* 16651 19839: contig of 3188 bp in length
* 19839 19859: gap of unknown length
* 19859 24632: contig of 4773 bp in length
* 24632 24651: gap of unknown length
* 24652 29640: contig of 4969 bp in length
* 29641 33877: contig of 4237 bp in length
* 33878 33897: gap of unknown length
* 33898 38499: contig of 4602 bp in length
* 38500 38519: gap of unknown length
* 38520 42610: contig of 4090 bp in length
* 42610 42629: gap of unknown length
* 42630 49217: contig of 6588 bp in length
* 49218 49237: gap of unknown length
* 49238 56851: contig of 7614 bp in length
* 56852 62381: contig of 5510 bp in length
* 62382 62401: gap of unknown length
* 62402 69660: contig of 7259 bp in length
* 69661 76728: contig of 7048 bp in length
* 76729 76748: gap of unknown length
* 76749 81905: contig of 5157 bp in length
* 81906 81925: gap of unknown length
* 81926 91698: contig of 9773 bp in length
* 91699 91718: gap of unknown length
* 91719 101313: contig of 9595 bp in length
* 101314 101334: gap of unknown length
* 101334 112260: contig of 10927 bp in length
* 112261 112280: gap of unknown length
```

```
/organism="proteobacteria"
/strain="PEST"
/db_xref="taxon:7165"
```

TITLE

TITLE
JOURNAL
Direct Submission
Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome
Research 320 Charles Street Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6558

Center clone name: 750_J_17

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1	846:	contig of 846 bp in length	20410	21222:	contig of 813 bp in length
*		gap of unknown length	*		gap of unknown length
*	847	contig of 819 bp in length	*	22040:	contig of 818 bp in length
*		gap of unknown length	*		gap of unknown length
*	1666	contig of 804 bp in length	*	22851:	contig of 811 bp in length
*		gap of unknown length	*		gap of unknown length
*	2470	contig of 829 bp in length	*	23667:	contig of 816 bp in length
*		gap of unknown length	*		gap of unknown length
*	3299	contig of 820 bp in length	*	24478:	contig of 811 bp in length
*		gap of unknown length	*		gap of unknown length
*	4119	contig of 810 bp in length	*	25290:	contig of 812 bp in length
*		gap of unknown length	*		gap of unknown length
*	4929	contig of 818 bp in length	*	26120:	contig of 830 bp in length
*		gap of unknown length	*		gap of unknown length
*	5747	contig of 807 bp in length	*	26967:	contig of 847 bp in length
*		gap of unknown length	*		gap of unknown length
*	6554	contig of 786 bp in length	*	27778:	contig of 811 bp in length
*		gap of unknown length	*		gap of unknown length
*	7340	contig of 841 bp in length	*	28568:	contig of 790 bp in length
*		gap of unknown length	*		gap of unknown length
*	8181	contig of 814 bp in length	*	29394:	contig of 826 bp in length
*		gap of unknown length	*		gap of unknown length
*	8995	contig of 799 bp in length	*	30213:	contig of 819 bp in length
*		gap of unknown length	*		gap of unknown length
*	9794	contig of 796 bp in length	*	31022:	contig of 809 bp in length
*		gap of unknown length	*		gap of unknown length
*	10590	contig of 811 bp in length	*	31833:	contig of 811 bp in length
*		gap of unknown length	*		gap of unknown length
*	11401	contig of 826 bp in length	*	32683:	contig of 850 bp in length
*		gap of unknown length	*		gap of unknown length
*	12227	contig of 809 bp in length	*	33528:	contig of 845 bp in length
*		gap of unknown length	*		gap of unknown length
*	13036	contig of 794 bp in length	*	34340:	contig of 812 bp in length
*		gap of unknown length	*		gap of unknown length
*	13830	contig of 801 bp in length	*	35161:	contig of 821 bp in length
*		gap of unknown length	*		gap of unknown length
*	14631	contig of 810 bp in length	*	35982:	contig of 821 bp in length
*		gap of unknown length	*		gap of unknown length
*	15441	contig of 854 bp in length	*	36796:	contig of 814 bp in length
*		gap of unknown length	*		gap of unknown length
*	16295	contig of 812 bp in length	*	37709:	contig of 913 bp in length
*		gap of unknown length	*		gap of unknown length
*	17107	contig of 816 bp in length	*	38535:	contig of 826 bp in length
*		gap of unknown length	*		gap of unknown length
*	17923	contig of 817 bp in length	*	39361:	contig of 826 bp in length
*		gap of unknown length	*		gap of unknown length
*	18740	contig of 771 bp in length	*	40182:	contig of 821 bp in length
*		gap of unknown length	*		gap of unknown length
*	19511	contig of 899 bp in length	*	41025:	contig of 843 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	41870:	contig of 845 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	42678:	contig of 808 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	43469:	contig of 791 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	44289:	contig of 820 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	45105:	contig of 816 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	45920:	contig of 815 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	46743:	contig of 823 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	47553:	contig of 810 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	48356:	contig of 803 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	49183:	contig of 827 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	50016:	contig of 833 bp in length
*		gap of unknown length	*		gap of unknown length
*		gap of unknown length	*	50820:	contig of 804 bp in length
*		gap of unknown length	*		gap of unknown length

Wed May 24 09:19:41 2000

```
* * *
* * 50821      gap of unknown length
* * 51621: contig of 801 bp in length
* * 51622      gap of unknown length
* * 52427: contig of 806 bp in length
* * 52428      gap of unknown length
* * 53234: contig of 807 bp in length
* * 53235      gap of unknown length
* * 54045: contig of 811 bp in length
* * 54046      gap of unknown length
* * 54866: contig of 821 bp in length
* * 54867      gap of unknown length
* * 55880: contig of 814 bp in length
* * 56508: contig of 828 bp in length
* * 55681      gap of unknown length
* * *

Query Match      10.4%; Score 46.6; DB 55; Length 63832;
Best Local Similarity 49.1%; Pred. No. 0.67;
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 218 aaaaaactaaagcagtagagaattacattacacagatggcagatatggacaactaagtg 277
      ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24988 AAAAAAAAAACACAAAAANANANACATAAAAAAAAAAAAAAAAAAAAAAAAAACA 24929

Qy 278 agaaggtatcagaacaaggtttaataagaatccttaaaaaagtaagccacaacacagaaa 337
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24928 AAAAAACAAAAAANANANANANANANANANANANANANANANANANANANANANAN 24869

Qy 338 agacaacaacagtgaaattcaacagaagaagtaattggactctgtgatgacgatt 397
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24868 ACNANANANANANANANANANANANANANANANANANANANANANANANANANAN 24809

Qy 398 attgaactacaagtgtctcacagactagaacttaacggaacaa 439
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24808 AAAAAAAAAAAAAAAAAAAAAAAAAACCAAAAAAAAAACAAAAA 24767
```

Search completed: May 24, 2000, 00:27:38
Job time: 12409 sec

Db 4536 AA 4595
QY 347 cagtgaattcaacagagaaagtaagctctgatgagatgacgattattgaacta 406
Db 4596 AA 4655
QY 407 caagtgtcacagactagaaacttaacggaacaa 439
Db 4656 AA 4688

RESULT 4
X33180
ID AC X33180 standard; DNA; 7797 BP.
AD X33180;
DT 25-JUN-1999 (first entry)
DE Cowpox virus bsr full length gene sequence.
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
OS Cowpox virus.
PN WO9913073-A2.
PD 18-MAR-1999.
PF 07-SEP-1998; J04010.
PR 08-SEP-1997; JP-259235.
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI Hamada H;
DR WPI; 99-243728/20.
PT New apoptosis-resistant virus-sensitive cell
PS Example 1; Page 34-38; 51pp; English.
CC The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the cowpox virus bsr gene which is used in an example from the present invention.
SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T;

Query Match 10.2%; Score 45.8; DB 1; Length 7797;
Best Local Similarity 48.0%; Pred. No. 0.0051;
Matches 131; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
QY 167 tggatcagtcgcccggccaggttaagtaacttagcactgttaagcctgaaacaaacta 226
Db 4841 TGACTGAGTCGCCGGGTACCGGTATCCAAATAAACCCCTTTGCAGTTGCAAAAAA 4900
QY 227 aagcagtagagaattacctatcacagatggcagaactaagtgagaaggtat 286
Db 4901 AA 4960
QY 287 cagaacaagggtttaagaaactccttaaaagtaagccaacagagaagacaacaa 346
Db 4961 AA 5020
QY 347 cagtgaattcaacagagaaagtaagctctgatgagatgacgattattgaacta 406
Db 5021 AA 5080

QY 407 caagtgtcacagactagaaacttaacggaacaa 439
Db 5081 AA 5113

RESULT 5
X33184
ID AC X33184 standard; DNA; 7996 BP.
AD X33184;
DT 25-JUN-1999 (first entry)
DE Base sequence of the plasmid pRx-Bcl 2-1-hcd 25.
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9913073-A2.
PD 18-MAR-1999.
PF 07-SEP-1998; J04010.
PR 08-SEP-1997; JP-259235.
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI Hamada H;
DR WPI; 99-243728/20.
PT New apoptosis-resistant virus-sensitive cell
PS Example 3; Page 46-49; 51pp; English.
CC The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the base sequence of the plasmid pRx-Bcl 2-1-hcd 25, which contains the human Bcl-2 gene, and is used in an example from the present invention.
SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T;

Query Match 10.2%; Score 45.8; DB 1; Length 7996;
Best Local Similarity 48.0%; Pred. No. 0.0051;
Matches 131; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
QY 167 tggatcagtcgcccggccaggttaagtaacttagcactgttaagcctgaaacaaacta 226
Db 5040 TGACTGAGTCGCCGGGTACCGGTATCCAAATAAACCCCTTTGCAGTTGCAAAAAA 5099
QY 227 aagcagtagagaattacctatcacagatggcagaactaagtgagaaggtat 286
Db 5100 AA 5159
QY 287 cagaacaagggtttaagaaactccttaaaagtaagccaacagagaagacaacaa 346
Db 5160 AA 5219
QY 347 cagtgaattcaacagagaaagtaagctctgatgagatgacgattattgaacta 406
Db 5220 AA 5279
QY 407 caagtgtcacagactagaaacttaacggaacaa 439

Db 5280 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5312

RESULT 6

T41852

ID T41852 standard; DNA: 9789 BP.

AC 20-FEB-1997 (first entry)

DE cDNA encoding Plasmodium falciparum erythrocyte membrane protein.

KW Plasmodium falciparum; erythrocyte membrane protein; malaria;

OS detection; identification; treatment; prevention; parasite; ss.

OS Plasmodium falciparum MC type.

FH Key Location/Qualifiers

FT 326..9497

FT /tag= a

FT /product= Erythrocyte membrane protein

FT misc_feature 518..520

FT /tag= b

FT /transl_except= GTA encodes Tyrosine

FT 656..658

FT /tag= c

FT /transl_except= ATT encodes Leucine

FT 2909..2911

FT /tag= d

FT /transl_except= AAC encodes Aspartic acid

FT 3461..3463

FT /tag= e

FT /transl_except= GAA encodes Glutamine

FT 5546..5548

FT /tag= f

FT /transl_except= CCT encodes Arginine

FT 6254..6256

FT /tag= g

FT /transl_except= AAT encodes Lysine

FT 6257..6259

FT /tag= h

FT /transl_except= ATA encodes Tyrosine

FT 6263..6265

FT /tag= i

FT /transl_except= AAC encodes Lysine

FT 6269..6271

FT /tag= j

FT /transl_except= TTC encodes Isoleucine

FT 6272..6274

FT /tag= k

FT /transl_except= ATA encodes Histidine

FT 6275..6277

FT /tag= l

FT /transl_except= ATT encodes Asparagine

FT 6278..6280

FT /tag= m

FT /transl_except= GGA encodes Tryptophan

FT 7754..8478

FT /tag= n

FT intron

FT W09633736-A1.

PD 31-OCT-1996.

PF 26-APR-1996; U05798.

PR 27-APR-1995; US-430908.

PA (AFFY-) AFFYMAX TECHNOLOGIES NV.

PI Baruch DI, Howard RU, Pasloske BL;

PI WPI: 96-497376/49.

DR P-PSDB; W00384.

PT New Plasmodium falciparum erythrocyte membrane proteins - used to

PT develop products for the diagnosis, treatment or prevention of

PT malaria parasite infections

PS Disclosure; Figure 12; 149pp; English.

CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte

CC membrane protein 1 (PfEMP1) or active fragments or analogues of that

CC protein can be used in the treatment or prevention of symptoms of a

CC malaria parasite infection. The polypeptides can inhibit, block or

CC reverse the sequestration of erythrocytes in patients suffering from

CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as

CC probes and primers to identify a Plasmodium falciparum parasite, the

CC primers used to generate characteristic amplification patterns from

CC different P. falciparum strains. Antibodies specifically

CC immunoreactive with the PfEMP1 polypeptide or its fragments may be

CC used in diagnosis of malaria infection. This sequence encodes the

CC PfEMP1 protein of the MC type of Plasmodium falciparum. An

CC alternative, truncated version of the coding sequence (a cDNA clone)

CC is given in T41853.

CC Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T;

Query Match 9.6%; Score 43; DB 1; Length 9789;

Best Local Similarity 49.8%; Pred. No. 0.031;

Matches 109; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 207 tgtaagcctgaaacaaactaaagcagatgacattacattacacagatgccaagatagg 266

DB 8122 TATATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8181

QY 267 acaactaagtgaaggtatcagaacaaaggttttaataagaaatccttaaaagtgtaagcca 326

DB 8182 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8241

QY 327 acaaacagaaagacaaacagtgaaattcaacagaaagaaagtaagtgaactctgatga 386

DB 8242 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8301

QY 387 agatgacgattattgaactacaagtgctcacagactaga 425

DB 8302 ATATTTTATTCATACATACATATACATATACATATATATA 8340

RESULT 7

V61487

ID V61487 standard; cDNA; 4237 BP.

AC V61487; 1999 (first entry)

DE Human secreted protein fe366_1 cDNA.

KW Secreted protein; human; fe366_1; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 3746..4030

FT /*tag= a

FT W09841539-A2.

PD 24-SEP-1998.

PF 19-MAR-1998; U05474.

PR 18-MAR-1998; US-040963.

PR 19-MAR-1997; US-820493.

PA (GENY) GENETICS INST INC.

PI Agostino WJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D,

PI Racine LA, Spaulding V, Treacy M;

PI WPI: 98-521163/44.

DR P-PSDB; W79096.

PT New polynucleotide(s) encoding secreted human proteins - derived

PT from human foetal kidney, adult testes and adult or foetal brain

PT cDNA libraries

PS Claim 36; Page 88-91; 112pp; English.

CC This full-length cDNA clone, designated fe366_1, codes for a novel

CC secreted human protein (see W79096). It was isolated from a human

CC adult brain cDNA library using methods which are selective for

CC cDNAs encoding secreted proteins, or was identified as encoding

CC a secreted or transmembrane protein on the basis of computer

CC analysis of the encoding protein. The nucleotide sequence shows

CC homology to some database sequences, and may contain a CAA repeat

CC and/or Alu repetitive element. The invention provides cDNA clones

CC (see V61477-87) from human foetal kidney, adult testis, and adult

CC or foetal brain cDNA libraries that code for secreted proteins

CC (see W79087-97). These clones are deposited as ATCC 98364. The

CC polynucleotides and proteins are predicted to have useful

CC biological activities which would make them suitable for treating,

CC preventing or ameliorating medical conditions in humans and

CC animals, although no supporting data is given. Suggested

CC activities include nutritional, immune stimulating (e.g. as

CC vaccines) or suppressing, haematopoiesis regulating, tissue growth,


```
Db 951 AGAAGAAAA 959
RESULT 12
V44592
ID V44592 standard; DNA; 6204 BP.
AC V44592;
DT 07-DEC-1998 (first entry)
DE Streptococcus pneumoniae IgA protease gene.
KW IgA protease; iga gene; pneumonia; meningitis; antibiotic; vaccine;
ss.
OS Streptococcus pneumoniae strain P110.
FH Key Location/Qualifiers
FT CDS 259..6153
   /tag= a
PN CA2196502-A.
PD 24-MAR-1998.
PF 31-JAN-1997; 196502.
PR 23-SEP-1996; US-026939.
PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.
PI (NEWE-) NEW ENGLAND MEDICAL CENT HOSPITALS INC.
PI Gilbert-Rothstein JV, Plaut AG, Weiser JN;
DR WPI; 98-39968/35.
DR P-PSDB; W69165.
PT New DNA encoding Streptococcus pneumoniae IgA protease .- and
   antibody screening assay and subunit vaccine etc
PS Claim 5; Fig 4; 66pp; English.
CC This isolated nucleic acid encodes Streptococcus pneumoniae IgA
   protease (see W69165). To isolate the gene, fragments of S.
   pneumoniae strain P110 genomic DNA were cloned into a shuttle
   vector, transformed into E. coli, and then used to transform S.
   pneumoniae strain R6x. Those DNA fragments found to induce changes
   in colony morphology were then selected for further study. The
   invention provides a method of identifying a candidate antibiotic
   compound that involves incubating the IgA protease with an IgA
   protein, and comparing IgA protease activity in the presence and
   absence of a test compound. S. pneumoniae cells or isolated IgA
   protease enzyme can be used in this method. The antibiotic is
   used to treat S. pneumoniae or other bacterial infections. The
   invention also relates to a subunit vaccine comprising an isolated
   IgA protease polypeptide. This can be used to prevent or reduce
   the severity of (S. pneumoniae) infections.
CC Sequence 6204 BP; 2221 A; 1143 C; 1253 G; 1587 T;
SQ
```

```
Query Match 8.7%; Score 39.2; DB 1; Length 6204;
Best Local Similarity 46.1%; Pred. No. 0.27;
Matches 131; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 130 gaagcagaatgagaacagtagtatttagcccaagttcttgatcagtcgcccgggcccagg 189
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1585 GAAGCAGTGTGTGACCAACAAAGGTGAGACTGAGCTTCAACACAGCTGCCAGATATCTGTG 1644
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 190 ttaagtaacttagcactgttaaagcctgaaaaaactaaagcagtagagaaattacccttata 249
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1645 GTAAGTGATAAAGGTGAACCCAGAGAGGTAGCCCATTCGCAGAAATATAAGGTAATATT 1704
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 250 cagatggcagaatattggacaactaagtggaaaggtatcagaacaaagtttaataagaaatc 309
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1705 GAGCAAGTAAACCTGAAACTCCGGTGTGAAGACCAACAAAGCAAGTCCAGCAAAAAA 1764
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 310 cttaaaaaagtagcacaacaaacagacaaacacacagtgaaattcaacagaagaaaa 369
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1765 GAACAGTTCCTATTAACCAACAGAAAGAACACACAGTAATCAATGAAGGTACTACA 1824
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 370 gtaatgacctctgaagatgacgattattgaactacaagtgc 413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1825 GAAGGACCTCAATTCAGAAGCAGAAAAATCCAGTTCACCTGC 1868
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
Q69159/c
ID Q69159 standard; cDNA to mRNA; 1611 BP.
```

```
AC Q69159;
DT 22-FEB-1995 (first entry)
DE Human Fc-alpha-R cDNA.
KW IgA Fc receptor; Fc-alpha-R; immunoglobulin; extracellular;
KW Intracellular; domain; soluble; vaccine; phagocytosis; HIV; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 40..903
   /tag= a
FT signal_peptide 40..102
   /product= Mammalian IgA Fc receptor
FT mat_peptide 103..900
   /tag= b
   /tag= c
PN US5328987-A.
PD 12-JUL-1994.
PF 05-JUL-1990; 548059.
PR 05-JUL-1990; US-548059.
PR 04-NOV-1992; US-971092.
PA (IMM ) IMMUNEX CORP.
PI Maliszewski CR;
DR WPI; 94-225324/27.
DR P-PSDB; R59920.
PT New human IgA Fc receptor and its soluble form - are useful as
   vaccine adjuvant and to prevent IgA-mediated phagocytosis, e.g.
   of HIV particles
PS Claim 1; Fig 1; 27pp; English.
CC This sequence encodes the human IgA Fc receptor protein (Fc-alpha-R).
   The mammalian IgA Fc receptor is a receptor protein which binds the
   immunoglobulin IgA. In its native configuration Fc-alpha-R consists
   of an extracellular domain which binds to IgA, a hydrophobic region
   which immobilises the protein within the cells lipid bilayer and an
   intracellular region. A soluble form of the receptor pref. contains
   residues 1-266 or 1-206 of the full length protein. The soluble
   form of the receptor may be used as a vaccine adjuvant to increase
   production of antigen-specific IgA, thus improving specific immunity,
   and to inhibit IgA-mediated phagocytosis. It inhibits the uptake by
   effector cells of IgA-coated HIV particles.
CC Sequence 1611 BP; 404 A; 395 C; 387 G; 425 T;
SQ
```

```
Query Match 8.7%; Score 39; DB 1; Length 1611;
Best Local Similarity 59.5%; Pred. No. 0.18;
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 264 tggacaactaagtgaaggtatcagaacaaaggtttaataagaaatccttaaaaaagtaag 323
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1512 TGGGCAACAGAGTGAGACTCTGTCTCAATAAAAAATAAAAAATAAAAAATAAAAAATA 1453
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 324 ccacaaacagaagaacacacacagtgaaattcaacagaagaaagtaat 374
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1452 TAAAAAAGAGAAGACATGAAGGAAGAATAAACAGGATATAAAAAAT 1402
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
Q80908
ID Q80908 standard; cDNA; 2223 BP.
AC Q80908;
DT 24-AUG-1995 (first entry)
DE Plasmodium falciparum ABRA gene cDNA.
KW Plasmodium falciparum ABRA gene; recombinant poxvirus;
KW multicomponent multistage malarial vaccines; immunogens;
KW malaria diagnosis; ss.
OS Plasmodium falciparum (pABRA-8).
FH Key Location/Qualifiers
FT CDS 1..2223
   /tag= a
PN WO9428930-A.
PD 22-DEC-1994.
PF 10-JUN-1994; 006652.
PR 11-JUN-1993; US-075783.
PR 09-JUN-1994; US-257073.
PA (VIRO-) VIROGENETICS CORP.
```

RESULT	15	
V41451		
ID	V41451 standard; cDNA; 202 BP.	
AC	V41451;	
DT	12-OCT-1998 (first entry)	
DE	Nucleotide sequence of the 3' portion of the BL229_22 clone.	
KW	Clone BL229_22; secreted protein; human cell; immuno-modulator; PCR;	
KW	anti-tumour agent; tissue growth promoter; haemostatic agent; primer;	
KW	thrombolytic agent; molecular weight marker; genetic disorder;	
KW	antibody; ds.	
OS	Homo sapiens.	
PN	W09821332-A2.	
PD	22-MAY-1998.	
PF	14-NOV-1997; U20740.	
PR	13-NOV-1997; US-969515.	
PR	15-NOV-1996; US-749745.	
PR	02-JUN-1997; US-867678.	
PA	(GENY) GENETICS INST INC.	
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,	
PI	Racie LA, Spaulding V, Treacy M;	
DR	WPI: 98-267936/26.	
PT	New nucleic acid encoding secreted protein from human cells -	
PT	potentially useful, e.g. as immuno-modulators, anti-tumour agents,	
PT	promoters of tissue growth, haemostatic and thrombolytic agents	
PS	Clam 33; Page 75; 98pp: English.	
CC	This is the nucleotide sequence of the clone BL229_22, used in the	
CC	method of the invention involving secreted proteins from human cells	
CC	potentially useful as immuno-modulators, anti-tumour agents, tissue	


```
Query Match      16.3%; Score 73.4; DB 1; Length 7218;
Best Local Similarity 6.1%; Pred. No. 6.4e-12;
Matches 26; Conservative 240; Mismatches 161; Indels 0; Gaps 0;

QY 23 cctggcgacgagcgttgagcgctgagagacagagcgtccgagctgcagcca 82
   |||
Db 1435 CRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1376

QY 83 aacacgggagtcgtgtagcgcccaacaggaagcaagcaggggaagcagaaatga 142
   |||
Db 1375 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316

QY 143 gaacagtagtcttagcccaagttctgtagcagtcgagccgagcttaagtaactag 202
   |||
Db 1315 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256

QY 203 cactgttaagcctgaacaaactaaagcagtagagaattaccttatacagatgycgaagat 262
   |||
Db 1255 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196

QY 263 atggacaactaagtgagaaggtatcagaacaaggtttaataagaatccttaaaaaagtaa 322
   |||
Db 1195 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136

QY 323 gccacaacacagaaaacacacagtagtaattcaacagagaagaaagtaatggactctg 382
   |||
Db 1135 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076

QY 383 atgaagatgacgattattgaactacaagtcagtcacagacagtagaacttaacggaacaagtc 442
   |||
Db 1075 RRRRRRRRRATCGCAAGCTCCCTCGACCTCGACGCAAGCTCGGAATTAATCTGTGAGCG 1016

QY 443 taggaca 449
Db 1015 TATGGCA 1009

RESULT 2
5231168-1
; Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 1
; LENGTH: 3095
5231168-1

Query Match      9.2%; Score 41.2; DB 7; Length 3095;
Best Local Similarity 48.0%; Pred. No. 0.0088;
Matches 118; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 204 acttgtaagcctgaaataaactaaagcagtagagaattaccttatacagatggcagaata 263
   |||
Db 1167 acttccagaagatgataaaaatgaaaagtgaacatgaatagtagaagttgaagaaat 1226

QY 264 tggacaactaagtgagaaggtatcagacaaggtttaataagaatccttaaaaaagtaag 323
   |||
Db 1227 tctaccagaagataaaaatgaaaaggtcaacatgaatagtagaggttgaaagaaattct 1286

QY 324 ccaacaacagaaagacaacacagtgaaattcaacagagaagaaagtaatgactctga 383
   |||
Db 1287 accgaagagatgataaaaatgaaaagttgacatgaatagtagaagttgaaagaaattct 1346

QY 384 tgaagatgacgattattgaactacaagtgtccacagactagaacttaacggaaagactct 443
   |||
Db 1347 accgaagataaaaatgaaaaggtcaacatgaatagtagaggttgaaagaaattctacc 1406

QY 444 aggaca 449
```

```
Db 1407 agaaga 1412
   |||

RESULT 3
US-08-874-186-42/c
; Sequence 42, Application US/08874186
; Patent No. 5989885
; GENERAL INFORMATION:
; APPLICANT: Teng, David H-F.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Perry III, William L.
; APPLICANT: Skolnick, Mark H.
; TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE KINASE
; TITLE OF INVENTION: 4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
; TITLE OF INVENTION: SUPPRESSOR IN VARIOUS TYPES OF CANCER
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/874,186
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,482
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Stephen A.
; REGISTRATION NUMBER: 38,609
; REFERENCE/DOCKET NUMBER: 24884-121392-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4848
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2799 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..764
; FEATURE:
; NAME/KEY: exon
; LOCATION: 765..892
; FEATURE:
; NAME/KEY: intron
; LOCATION: 893..2799
US-08-874-186-42

Query Match      9.1%; Score 41; DB 4; Length 2799;
Best Local Similarity 58.7%; Pred. No. 0.0096;
Matches 71; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 269 actaagtgagaggttatcagacaaggttttaataagaaatccttaaaaaagtaagccaac 328
   |||
Db 1198 AACTTTCTGAGAAGTTACCATCTGTTTAATGAGTTGAGTCAATATATGAAAAACAAT 1139

QY 329 aaacagaaaagacaacacacagtgaaattcaacagagaaaaagtaattggactctgatgaag 388
```

Db 1138 AATCAGAGAGACCAACTGCTAAATACCAAGTAAGAGAGAAATAAGCAAGACAATGAAA 1079
QY 389 a 389
Db 1078 A 1078
RESULT 4
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,075
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95238
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3770..4013
; OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from
; plasmid RP4"
US-09-056-075-1
Query Match 9.0%; Score 40.4; DB 4; Length 6243;
Best Local Similarity 47.3%; Pred. No. 0.021;
Matches 122; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 116 aagcaagcagcagggaagcagcaaatgagaaacagatctttagcccaagttcttgatcagt 175
Db 1137 AAAAAAGACAGGTACAAAGTACCCCTATATACACGCTGAAAAATGAGGTAAAAAT 1196
QY 176 cggccggccaggttaagtaacttagcactgttaagcctgaaactaaagcagtag 235
Db 1197 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1256
QY 236 agaattacctatcacagatggcgaagatatggacaactaagtggagaaggtatcagaacaag 295
Db 1257 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1316

QY 296 gttaaatagaaatccttaaaaaagtaagccaaacagaaacagaaacagaaacagaaat 355
Db 1317 TATAAAATAAAAAATATAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1376
QY 356 tcaacagagaagaagaagtaa 373
Db 1377 AAAAAATAAAAAATAAAAA 1394
RESULT 5
US-08-790-912-1
; Sequence 1, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-4U1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6201
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1..30, 34..39, 46..51, 55..99, 103..135, 139
; LOCATION: 153, 157..165, 169..186, 190..210, 217..225,
; LOCATION: 229..234, 238..6150, 6154..6177, 6181..6201)
US-08-790-912-1
Query Match 8.7%; Score 39.2; DB 4; Length 6201;
Best Local Similarity 46.1%; Pred. No. 0.046;
Matches 131; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 130 gaagcagaaatgagaaacagatctttagcccaagttcttgatcagtcggccggccag 189
Db 1585 GAACGACTTGTGACCGACAAAGGTGAGACTGAGGTTCAACACGAGATGCCAGATACTGTG 1644

Db 1452 TAAAAAGAGAGAACAACTGAAGGAAGAAATTAACAGGATATAAAAAT 1402

QY 190 ttaagtaacttagcactgttaagcctgtaaaagcctgtaaaagcagtagagaaattacccttata 249
Db 1645 GTAAGTGATAAAGGCTGAACACGAGCAGGTAGCCCATTCGCCAGATATAAAGGGTAATATT 1704
QY 250 cagatgcaagatatggacaactaagtgaaggttatcagacaaggttttaataagaatc 309
Db 1705 GAGCAAGTAAACCTGAACTCGGTTGAGAAGACCAAGAAACAAGGTCAGAAAAAACT 1764
QY 310 cttaaaaaagtagcagcaaacacagaaagacacaaacagtagtgaattcaacagagaagaaa 369
Db 1765 GAAGAAGTTCCAGTAAACACACAGAAAGAACACACAGTAAATCCAAATGAAGGTACTACA 1824
QY 370 gtaatgactctgtaagatgacgattattgaactacaaagtgc 413
Db 1825 GAAGGAACCTCAATTCAGAAGCAGAAATCCAGTTCAACCTGC 1868

RESULT 6
US-07-971-092-1/c
: Sequence 1, Application US/07971092
: Patent No. 5328987
: GENERAL INFORMATION:
: APPLICANT: Maliszewski, Charles R.
: TITLE OF INVENTION: Huiga Fc Receptor
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Immunex
: STREET: 51 University
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98101
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/971,092
: FILING DATE: 19921104
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Perkins, Patricia A.
: REGISTRATION NUMBER: 34693
: REFERENCE/DOCKET NUMBER: 2603
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1589 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Human IgA Fc Receptor
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 40..900
US-07-971-092-1

Query Match 8.7%; Score 39; DB 1; Length 1589;
Best Local Similarity 59.5%; Pred. No. 0.028;
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 264 tggacaactaagtgaaggttatcagacaaggttttaataagaatcccttaaaaaagtaag 323
Db 1512 TGGGCAACAGAGTGAGACTCTCTCTCAATAAAAAATAAAATAAAATAAAATAAA 1453
QY 324 ccaacaacagaaagacacacacagtagtgaattcaacagagaagaaagtaat 374

US-08-257-073-4
: Sequence 4, Application US/08257073
: Patent No. 5766597
: GENERAL INFORMATION:
: APPLICANT: Paoletti, Enzo
: APPLICANT: de Taisne, Charles
: APPLICANT: Tine, John A.
: TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Curtis, Morris & Safford, P.C.
: STREET: 530 Fifth Avenue, 25th Floor
: CITY: New York
: STATE: New York
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/257,073
: FILING DATE: 09-JUN-1994
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/075,783
: FILING DATE: 11-JUN-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/852,305
: FILING DATE: 18-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,183
: FILING DATE: 20-MAR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Frommer, William S.
: REGISTRATION NUMBER: 25,506
: REFERENCE/DOCKET NUMBER: 454310-2570
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 840-3333
: TELEFAX: (212) 840-0712
: TELEX: 425066 CURTMS
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2223 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-257-073-4

Query Match 8.7%; Score 39; DB 2; Length 2223;
Best Local Similarity 47.4%; Pred. No. 0.033;
Matches 117; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
QY 201 agcactgttaagcctgaaaaaactaaagcagtagagaattacccttatcacagatggcaag 260
Db 1968 AGCAATCACCAGCAAAATAGAAGCTGAGTGGATGCCCTCGCACCAAAAATAAGGAAGA 2077
QY 261 atatgacaactaagtgaaggttatcagacaaggttttaataagaatcccttaaaaaagt 320
Db 2028 AGAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAA 2087
QY 321 aagcccaacacagaaagacacacacagtgaaattcaacagagaagaaagtaagttaattgc 380
Db 2088 AAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAA 2147
QY 381 tgatgaagtagcagattattgaactacaagtagctcacagactagaacttaacggaacag 440


```

RESULT 10
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; ADDRESSEE: DEFENSE COMMAND
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,417
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BIFFONI, ULYSSES J
; REGISTRATION NUMBER: 39,908
; REFERENCE/DOCKET NUMBER: DAM 398-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 410-671-1158
; TELEFAX: 410-671-2534
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-628-417-6

```

```

Query Match      8.4%; Score 38; DB 1; Length 240;
Best Local Similarity 50.5%; Pred. No. 0.022;
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 192 aagtaacttagcacttgtaagcctgaaaaactaaagcagtagagataacacctataca 251
Db 27 AAATTAATTTTACTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 86

Qy 252 gatggcaagatatggacaactaagtgaaggtatcgacaagaaggtttaatagaatcct 311
Db 87 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 146

Qy 312 taaaaaagtagcacaacaaacagaaagacacacagtgaaattcaacagagaagaagt 371
Db 147 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 206

Qy 372 aa 373
Db 207 AA 208

RESULT 11
5198342-1/c
; Patent No. 5198342
; APPLICANT: MALIIZEWSKI, CHARLES R.
; TITLE OF INVENTION: DNA ENCODING IGA FC RECEPTORS
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/07/548,059
; FILING DATE: 05-JUL-1990
; SEQ ID NO:1
; LENGTH: 1611
5198342-1

Query Match      8.3%; Score 37.4; DB 7; Length 1611;
Best Local Similarity 58.6%; Pred. No. 0.081;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 264 tggacaactaagtgaaggtatcgacaagaaggttttaatagaaatccttaaaaaagtaag 323
Db 1512 TGGCAACAGAGTCGAGACTCTGTCTCAATAAATAAATAAATAAATAAATAAATAA 1453

Qy 324 ccaacaaacagaaaagacaacacagcagtgaaattccaacagagaagaagtaat 374
Db 1452 TAAAAAAGAGAGAACAATGAAGGAAGAATAATAACACGGATATAAANAAT 1402

RESULT 12
US-08-520-678A-24/c
; Sequence 24, Application US/08520678A
; Patent No. 5874565
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles M.
; APPLICANT: Kolykhalov, Alexander A.
; TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
; TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howell & Haferkamp, L.C.
; STREET: 7733 Forsyth Blvd., Suite 1400
; CITY: St. Louis
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/520,678A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Henderson, Melodie W.
; REGISTRATION NUMBER: 37,848
; REFERENCE/DOCKET NUMBER: 6029-6836
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; TELEX:
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-520-678A-24

```

```

Query Match      8.2%; Score 37; DB 3; Length 257;
Best Local Similarity 54.9%; Pred. No. 0.044;
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 253 atggcaagatatgacaactaagtgaaggtatcgacaagaaggtttaatagaatcctt 312
Db 241 AGGCTAAGATGGAGCCACCCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 182

Qy 313 aaaaaagtaagccaacaaacagaaagacacacagtgaaattccaacagagaagaagta 372

```


Search completed: May 23, 2000, 23:55:11
Job time: 10423 sec

	Query Match	8.0%;	Score 36;	DB 4;	Length 1512;
	Best Local Similarity	53.6%;	Prod. NO. 0.2;		
	Matches 75;	Conservative	0;	Mismatches 65;	Indels 0;
	Gaps	0;			
QY	215	ctgaaaaaactaaacagcagtagaataattacctttatacagatggcagatattggaacactaa	274		
Db	1373	CTGAATTCCTTCAAAATACCAAAATATTTTTTCACAAAATTTAGAAAAAATTCACAAA	1432		
QY	275	qtgaagaagtatcagaacacaggtttatagaataatccttaaaaaagtaagccaaacacag	334		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:01:42 ; Search time 3550.87 Seconds
(without alignments)
-528.346 Million cell updates/sec

Title: US-09-215-435-116
Perfect score: 450
Sequence: 1 ctgctccagcgtgacgcg.....acggacaagtctaggacag 450

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues

Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA:*

- 1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
- 2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
- 3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
- 4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
- 5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq:*
- 6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq:*
- 7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq:*
- 8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq:*
- 9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq:*
- 10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq:*
- 11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq:*
- 12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq:*
- 13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq:*
- 14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq:*
- 15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq:*
- 16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq:*
- 17: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
- 18: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
- 19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq:*
- 20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq:*
- 21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq:*
- 22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq:*
- 23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq:*
- 24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq:*
- 25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq:*
- 26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq:*
- 27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq:*
- 28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq:*
- 29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq:*
- 30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq:*
- 31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq:*
- 32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq:*
- 33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq:*
- 34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq:*
- 35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq:*
- 36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq:*
- 37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq:*
- 38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq:*
- 39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq:*
- 40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq:*
- 41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq:*
- 42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq:*
- 43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq:*
- 44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq:*

- 45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq:*
- 46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq:*
- 47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq:*
- 48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq:*
- 49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq:*
- 50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq:*
- 51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq:*
- 52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq:*
- 53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq:*
- 54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq:*
- 55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq:*
- 56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq:*
- 57: /cgn2_6/ptodata/1/pna/US095_COMB.seq:*
- 58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
- 59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq:*
- 60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq:*
- 61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq:*
- 62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq:*
- 63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq:*
- 64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq:*
- 65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq:*
- 66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq:*
- 67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq:*
- 68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq:*
- 69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
- 70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
- 71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
- 72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq:*
- 73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq:*
- 74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq:*
- 75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq:*
- 76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq:*
- 77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq:*
- 78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq:*
- 79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq:*
- 80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq:*
- 81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq:*
- 82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq:*
- 83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq:*
- 84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq:*
- 85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq:*
- 86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq:*
- 87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq:*
- 88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
- 89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq:*
- 90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq:*
- 91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq:*
- 92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq:*
- 93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq:*
- 94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq:*
- 95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq:*
- 96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq:*
- 97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq:*
- 98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
- 99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
- 100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
- 101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
- 102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
- 103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	IDB	Description
1	450	100.0	450	43	US-09-215-435-116 Sequence 116, App
2	450	100.0	548	90	US-60-164-285-6488 Sequence 6488, Ap
3	450	100.0	559	43	US-09-215-435-516 Sequence 516, App

Wed May 24 09:19:42 2000

us-09-215-435-116.rnp

4 450 100.0 568 46 US-09-289-768-27719
5 450 100.0 568 49 US-09-332-782-28739
6 450 100.0 577 25 US-08-870-870-65
7 450 100.0 577 25 US-08-870-870A-65
8 450 100.0 589 92 US-60-172-373-13544
9 450 100.0 589 92 US-09-359-922-4808
c 10 439 97.6 590 51 US-09-359-922-4808
11 384.2 85.4 421 55 US-09-431-517-13611
12 384.2 85.4 421 55 US-09-431-517-13612
13 372 82.7 420 55 US-09-431-517-13610
14 362 80.4 476 52 US-09-362-510-58828
15 361 80.2 626 92 US-60-172-373-25470
16 354.4 78.8 393 49 US-09-332-782-16283
17 354.4 78.8 393 57 US-09-515-694-16283
18 348.4 77.4 519 45 US-09-277-227-18621
19 348.4 77.4 519 50 US-09-346-956-17502
20 342.8 76.2 347 26 US-08-905-135-285
21 340.8 75.7 471 46 US-09-287-618-25346
c 22 337 74.9 338 49 US-09-332-782-6058
23 337 74.9 338 57 US-09-515-694-6058
24 302.2 67.2 407 52 US-09-362-510-29313
25 298.6 66.4 726 94 US-60-177-646-328
26 292 64.9 314 23 US-08-826-438-1946
27 292 64.9 314 30 US-08-951-197-1946
28 292 64.9 314 59 US-60-013-696-6156
29 291 64.7 306 19 US-08-706-785-996
30 283 62.9 312 36 US-09-021-031-30
31 283 62.9 312 65 US-60-039-325-30
32 280 62.2 294 22 US-08-813-561-30
33 280 62.2 294 59 US-60-013-062-30
34 273.4 60.8 376 45 US-09-271-490-958
35 270 60.0 270 24 US-08-845-751-894
36 269.8 60.0 283 5 US-08-104-507A-135
37 269.8 60.0 283 5 US-08-104-507C-135
38 269.4 59.9 351 3 US-07-916-491B-385
39 269.4 59.9 351 3 US-07-977-780D-443
40 269.4 59.9 351 5 US-08-100-523B-440
41 264.6 58.8 408 3 US-07-977-780D-2279
42 264.6 58.8 408 5 US-08-100-523B-2244
43 264 58.7 267 33 US-08-986-693-3640
44 264 58.7 267 33 US-08-986-693A-3640
45 262 58.2 264 34 US-08-999-861-3451

ALIGNMENTS

RESULT 1
US-09-215-435-116
; Sequence 116, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215.435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 116
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

Sequence 27719, A
Sequence 28739, A
Sequence 65, App1
Sequence 13544, A
Sequence 4808, Ap
Sequence 4808, Ap
Sequence 13611, A
Sequence 14582, A
Sequence 13610, A
Sequence 58828, A
Sequence 25470, A
Sequence 16283, A
Sequence 16283, A
Sequence 18621, A
Sequence 17502, A
Sequence 285, App
Sequence 25346, A
Sequence 6058, Ap
Sequence 6058, Ap
Sequence 2313, A
Sequence 328, App
Sequence 1946, Ap
Sequence 1946, Ap
Sequence 6156, Ap
Sequence 996, App
Sequence 30, App1
Sequence 30, App1
Sequence 30, App1
Sequence 958, App
Sequence 894, App
Sequence 135, App
Sequence 135, App
Sequence 385, App
Sequence 443, App
Sequence 440, App
Sequence 2279, Ap
Sequence 2244, Ap
Sequence 3640, Ap
Sequence 3640, Ap
Sequence 3451, Ap

; NAME/KEY: CDS
; LOCATION: 25...399
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 25..186
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 3.5
; OTHER INFORMATION: seq SIIAQVLDQSARA/RL
US-09-215-435-116

Query Match 100.0%; Score 450; DB 43; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.5e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgtccagcctgaacccagccatgagcagagagcttgagcgctgagagacag 60
Db 1 ctgtccagcctgacgcccagccatgagcagagagcttgagcgctgagagacag 60

QY 61 aggtggcgcagctgacgcccagccaaacacacggggtatcctggtgatcgcccaacaggaagca 120
Db 61 aggtggcgcagctgacgcccagccaaacacacggggtatcctggtgatcgcccaacaggaagca 120

QY 121 aagcacaggggaagcagaataagaaacagtatcttagcccaagttctgagcagtcgccc 180
Db 121 aagcacaggggaagcagaataagaaacagtatcttagcccaagttctgagcagtcgccc 180

QY 181 cgggcccaggttaagtaacttagcacttgaagcctgaaacactgaaaaactaaagcagtagagaat 240
Db 181 cgggcccaggttaagtaacttagcacttgaagcctgaaacactgaaaaactaaagcagtagagaat 240

QY 241 tacctatcacagatggcaagatatggcaactaagtgagaaggtatcagaacaagggttta 300
Db 241 tacctatcacagatggcaagatatggcaactaagtgagaaggtatcagaacaagggttta 300

QY 301 atagaaatccttaaaagtagcagccaaacacagaaagacaaacacagtgaaattcaac 360
Db 301 atagaaatccttaaaagtagcagccaaacacagaaagacaaacacagtgaaattcaac 360

QY 361 agaagaaaagtaagtgactctgtagaagtagcagattattgaactacaagtgctcacaga 420
Db 361 agaagaaaagtaagtgactctgtagaagtagcagattattgaactacaagtgctcacaga 420

QY 421 ctagaacttaacggaaacaaagtcaggacag 450
Db 421 ctagaacttaacggaaacaaagtcaggacag 450

RESULT 2

US-60-164-285-6488
; Sequence 6488, Application US/60164285
; GENERAL INFORMATION:
; APPLICANT: Ma, Xiao-Jun
; TITLE OF INVENTION: Tumor Associated Molecules (TAMs): Targets for diagnosis, trea
; FILE REFERENCE: 3214
; CURRENT APPLICATION NUMBER: US/60/164,285
; CURRENT FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 8259
; SEQ ID NO 6488
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-164-285-6488

Query Match 100.0%; Score 450; DB 90; Length 548;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgtccagcctgacgcccagccatgagcagagagcttgagcgctgagagacag 60
Db 15 ctgtccagcctgacgcccagccatgagcagagagcttgagcgctgagagacag 74

```
QY 61 aggtgcccagctgagcggccaaacacgaggatccttggtgatcgccccaacaggagca 120
|||
Db 75 aggtgcccagctgagcggccaaacacgaggatccttggtgatcgccccaacaggagca 134
|||
QY 121 aagcacagggaagcagaaatgagaacacagtatcttagcccaagttctggtatcggtcc 180
|||
Db 135 aagcacagggaagcagaaatgagaacacagtatcttagcccaagttctggtatcggtcc 194
|||
QY 181 cgggcccaggttaagttaacttagcactgttaagcctgaaacaaactaaagcagtagagaat 240
|||
Db 195 cgggcccaggttaagttaacttagcactgttaagcctgaaacaaactaaagcagtagagaat 254
|||
QY 241 tacctatcacagatgccaagatatgacaactaagtgaagaagtatcagaacaagttta 300
|||
Db 255 tacctatcacagatgccaagatatgacaactaagtgaagaagtatcagaacaagttta 314
|||
QY 301 atagaaatccttaaaaaagtaagccaaacaaagaaagaaacacacagtgaaattcaac 360
|||
Db 315 atagaaatccttaaaaaagtaagccaaacaaagaaagaaagaaacacacagtgaaattcaac 374
|||
QY 361 agaagaaaagtaagtactgtactgtgaagatgaagatgacgattattgaaactacaagtgtcacaga 420
|||
Db 375 agaagaaaagtaagtactgtactgtgaagatgaagatgacgattattgaaactacaagtgtcacaga 434
|||
QY 421 ctagaacttaacggaacaaagtctaggacag 450
|||
Db 435 ctagaacttaacggaacaaagtctaggacag 464
|||
```

```
RESULT 3
US-09-215-435-516
; Sequence 516, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 516
; LENGTH: 559
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-215-435-516
```

```
Query Match 100.0%; Score 450; DB 43; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ctgctccagcgtgacgcccagccatggcgcagcagagagcgttaggcgctgagagacag 60
|||
Db 1 ctgctccagcgtgacgcccagccatggcgcagcagagagcgttaggcgctgagagacag 60
|||
QY 61 aggtgcccagctgagcggccaaacacgaggatccttggtgatcgccccaacaggagca 120
|||
Db 61 aggtgcccagctgagcggccaaacacgaggatccttggtgatcgccccaacaggagca 120
|||
QY 121 aagcacagggaagcagaaatgagaacacagtatcttagcccaagttctggtatcggtcc 180
|||
Db 121 aagcacagggaagcagaaatgagaacacagtatcttagcccaagttctggtatcggtcc 180
|||
```

```
QY 181 cgggcccaggttaagttaacttagcactgttaagcctgaaacaaactaaagcagtagagaat 240
|||
Db 181 cgggcccaggttaagttaacttagcactgttaagcctgaaacaaactaaagcagtagagaat 240
|||
QY 241 tacctatcacagatgccaagatatgacaactaagtgaagaagtatcagaacaagttta 300
|||
Db 241 tacctatcacagatgccaagatatgacaactaagtgaagaagtatcagaacaagttta 300
|||
QY 301 atagaaatccttaaaaaagtaagccaaacaaagaaagaaacacacagtgaaattcaac 360
|||
Db 301 atagaaatccttaaaaaagtaagccaaacaaagaaagaaagaaacacacagtgaaattcaac 360
|||
QY 361 agaagaaaagtaagtactgtactgtgaagatgaagatgacgattattgaaactacaagtgtcacaga 420
|||
Db 361 agaagaaaagtaagtactgtactgtgaagatgaagatgacgattattgaaactacaagtgtcacaga 420
|||
QY 421 ctagaacttaacggaacaaagtctaggacag 450
|||
Db 421 ctagaacttaacggaacaaagtctaggacag 450
|||
```

```
Query Match 100.0%; Score 450; DB 46; Length 568;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ctgctccagcgtgacgcccagccatggcgcagcagagagcgttaggcgctgagagacag 60
|||
Db 36 ctgctccagcgtgacgcccagccatggcgcagcagagagcgttaggcgctgagagacag 95
|||
QY 61 aggtgcccagctgagcggccaaacacgaggatccttggtgatcgccccaacaggagca 120
|||
Db 96 aggtgcccagctgagcggccaaacacgaggatccttggtgatcgccccaacaggagca 155
|||
QY 121 aagcacagggaagcagaaatgagaacacagtatcttagcccaagttctggtatcggtcc 180
|||
Db 156 aagcacagggaagcagaaatgagaacacagtatcttagcccaagttctggtatcggtcc 215
|||
QY 181 cgggcccaggttaagttaacttagcactgttaagcctgaaacaaactaaagcagtagagaat 240
|||
Db 216 cgggcccaggttaagttaacttagcactgttaagcctgaaacaaactaaagcagtagagaat 275
|||
QY 241 tacctatcacagatgccaagatatgacaactaagtgaagaagtatcagaacaagttta 300
|||
Db 276 tacctatcacagatgccaagatatgacaactaagtgaagaagtatcagaacaagttta 335
|||
QY 301 atagaaatccttaaaaaagtaagccaaacaaagaaagaaacacacagtgaaattcaac 360
|||
Db 336 atagaaatccttaaaaaagtaagccaaacaaagaaagaaagaaacacacagtgaaattcaac 395
|||
QY 361 agaagaaaagtaagtactgtactgtgaagatgacgattattgaaactacaagtgtcacaga 420
|||
```

Wed May 24 09:19:42 2000

us-09-215-435-116.rnp

```
Db 396 agaagaaagtaagtgactctgatgaagatgacgattattgaactacaagtgcacaga 455
QY 421 ctagaacttaacggaacaagtctaggacag 450
Db 456 ctagaacttaacggaacaagtctaggacag 485

RESULT 5
US-09-332-782-28739
; Sequence 28739, Application US/09332782A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/332,782A
; CURRENT FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28739
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(568)
; OTHER INFORMATION: n = A,T,C or G
US-09-332-782-28739
```

```
Query Match 100.0%; Score 450; DB 49; Length 568;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgtctcagcgctgacgcccagccatggcgagcagagagcttgaggcgtgaggagacag 60
Db 36 ctgtctcagcgctgacgcccagccatggcgagcagagagcttgaggcgtgaggagacag 95
QY 61 aggtgtgcccagcgctgacgcccagccatggcgagcagagagcttgaggcgtgaggagacag 120
Db 96 aggtgtgcccagcgctgacgcccagccatggcgagcagagagcttgaggcgtgaggagacag 155
QY 121 aagcacagggagcagacaataatgagaaacagatctttagccaaagtcttgatcagtcggcc 180
Db 156 aagcacagggagcagacaataatgagaaacagatctttagccaaagtcttgatcagtcggcc 215
QY 181 cgggcccaggttaagttaacttagcacttgtaagcctgaaacaaactaaagcagtagagaat 240
Db 216 cgggcccaggttaagttaacttagcacttgtaagcctgaaacaaactaaagcagtagagaat 275
QY 241 taccttatacagatggcaagatggaacataagtgagaaggttatcagaaagggttta 300
Db 276 taccttatacagatggcaagatggaacataagtgagaaggttatcagaaagggttta 335
QY 301 atagaaatccttaaaaagttaagccaaacagaaagaaacaaacaaacagtgaaattcaac 360
Db 336 atagaaatccttaaaaagttaagccaaacagaaagaaacaaacaaacagtgaaattcaac 395
QY 361 agaagaaagtaagtgaactctgatgaagatgacgattattgaactacaagtgcacaga 420
Db 396 agaagaaagtaagtgaactctgatgaagatgacgattattgaactacaagtgcacaga 455
QY 421 ctagaacttaacggaacaagtctaggacag 450
Db 456 ctagaacttaacggaacaagtctaggacag 485
```

```
RESULT 6
US-08-870-65
; Sequence 65, Application US/08870870
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
```

```
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,870
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0300 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITUT02
; CLONE: 2095728
; US-08-870-870-65
```

```
Query Match 100.0%; Score 450; DB 25; Length 577;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgtctcagcgctgacgcccagccatggcgagcagagagcttgaggcgtgaggagacag 60
Db 48 ctgtctcagcgctgacgcccagccatggcgagcagagagcttgaggcgtgaggagacag 107
QY 61 aggtgtgcccagcgctgacgcccagccatggcgagcagagagcttgaggcgtgaggagacag 120
Db 108 AGGCTGCCGCGAGCTGCAGGCCAAACACACAGGGGATCTCTGGTGATCGCGGCCACACAGGAACA 167
QY 121 aagcacagggagcagacaataatgagaaacagatctttagccaaagtcttgatcagtcggcc 180
Db 168 AAGCACAGGGGAAGCAGAAATAGAAACACATCTTATAGCCCAAGTCTTGGATCAGTCGGCC 227
QY 181 cgggcccaggttaagttaacttagcacttgaaagcctgaaacaaactaaagcagtagagaat 240
Db 228 CGGCGCCAGGTTAAGTAACTTAGCAGCTTGTAAAGCCCTGAAAAAATAAGCAGTAGAGAAT 287
QY 241 taccttatacagatggcaagatggaacataagtgagaaggttatcagaaagggttta 300
Db 288 TACCTTATACAGATGGCAAGATATGGACAACTAAGTGAGAAGGTATCAGAACAAAGTTTA 347
QY 301 atagaaatccttaaaaagttaagccaaacagaaagaaacaaacagtcgaaattcaac 360
Db 348 ATAGAAATCCTTTAAAAAAGTAGCCCAACAAACAGAGAAAGACACACACAGTGAATTTCAAC 407
```

Qy 361 agagaaaagtaagtgactctgatgaagatgacgattattgaactacaagtgtctcacaga 420
|||||
Db 408 AGAGAAAAGTAAATGGACTCTGATGAAGATGACGATTAATGAAGTGTCTCACAGA 467
|||||
Qy 421 ctagaacttaacggaacaagtctaggacag 450
|||||
Db 468 CTAGAAGTTAAGCGGAACAAGTCTAGGACAG 497
|||||

RESULT 7
US-08-870-870A-65
; Sequence 65, Application US/08870870A
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,870A
; FILING DATE: June 6, 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0300 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITUT02
; CLONE: 2095728
US-08-870-870A-65

Query Match 100.0%; Score 450; DB 25; Length 577;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgtctccagcgtgacgcccagccatgctgacagaggagcttgaggcgtgaggagacag 60
|||||
Db 48 CTGCTCCAGCGCTGACGCCGAGCCATGGCGGACGAGGAGCTTGAGGCGCTGAGGAGACAG 107
|||||
Qy 61 aggtctggccgagctgaagcccaacacacaggggatcctgtgtgatcgccccaacaggaaaga 120
|||||
Db 108 AGGCTGGCCGAGCTGACGGCCCAACACAGGGGATCCTGTGTGATGCGGGCCCAACAGGAAGA 167
|||||
Qy 121 aagcacagggagcagagaatgagaacacagtattcttagcccaagttctgagtcagtcgccc 180
|||||

Db 168 AAGCACAGGGAAGCAGAAAAATGAGAAACACAGTATCTTAGCCCAAGTTCTGTGATCAGTCGGCC 227
Qy 181 cgggccaggttaagtaacttagcactttaaagcctgaaaaaactaaagcagtagagaat 240
|||||
Db 228 CGGGCCAGGTTAAGTAACTTAGCAGCTTGTAAAGCCTGAAAAAATTAAGCAGTAGAGAAAT 287
|||||
Qy 241 taccttatcacagatggcgaagatatgacaaactaaagtgaaggttatcagaacaaggttta 300
|||||
Db 288 TACCTTATACAGATGGCAAGATATGGACAACTAAGTGAAGGTATCAGAACAAAGTTTA 347
|||||
Qy 301 atagaaatccttaaaaaagtaagcccaacaacagaaaaagacaacaacagtgaattcaac 360
|||||
Db 348 ATAGAAATCCTTAAAAAAGTAAGCCCAACAAACAGAAAAAGACAACACAGTGAATTTCAAC 407
|||||
Qy 361 agaagaaaagtaagtgactctgatgaagatgacgattattgaactacaagtgtctcacaga 420
|||||
Db 408 AGAGAAAAGTAAATGGACTCTGATGAAGATGACGATTAATGAAGTGTCTCACAGA 467
|||||
Qy 421 ctagaacttaacggaacaagtctaggacag 450
|||||
Db 468 CTAGAAGTTAAGCGGAACAAGTCTAGGACAG 497
|||||

RESULT 8
US-60-172-373-13544
; Sequence 13544, Application US/60172373
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym
; FILE REFERENCE: GX-0006 P
; CURRENT APPLICATION NUMBER: US/60/172,373
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 25,772
; SOFTWARE: PERL Program
; SEQ ID NO 13544
; LENGTH: 589
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 235725.6
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 8
; OTHER INFORMATION: a, t, c, g, or other
US-60-172-373-13544

Query Match 100.0%; Score 450; DB 92; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgtctccagcgtgacgcccagccatgctgacagaggagcttgaggcgtgaggagacag 60
|||||
Db 60 ctgtctccagcgtgacgcccagccatgctgacagaggagcttgaggcgtgaggagacag 119
|||||
Qy 61 aggtctggccgagctgaagcccaacacacaggggatcctgtgtgatcgccccaacaggaaaga 120
|||||
Db 120 aggtctggccgagctgaagcccaacacacaggggatcctgtgtgatcgccccaacaggaaaga 179
|||||
Qy 121 aagcacagggagcagagaatgagaacacagtattcttagcccaagttctgagtcagtcgccc 180
|||||
Db 180 aagcacagggagcagagaatgagaacacagtattcttagcccaagttctgagtcagtcgccc 239
|||||
Qy 181 cgggccaggttaagtaacttagcactttaaagcctgaaaaaactaaagcagtagagaat 240
|||||
Db 240 cgggccaggttaagtaacttagcactttaaagcctgaaaaaactaaagcagtagagaat 299
|||||
Qy 241 taccttatcacagatggcgaagatatgacaaactaaagtgaaggttatcagaacaaggttta 300
|||||

Db 300 tacctatagatggaagatggaacactaagtgaaggtatcagaacaagggtta 359
Qy 301 atagaatactttaaaagctaaagcaacaaacagaaagacaacacagtgaaattcaac 360
Db 360 atagaatactttaaaagctaaagcaacaaacagaaagacaacacagtgaaattcaac 419
Qy 361 agagaataatgaatgactctgatgaatgacgatattattgaactacaagtgtcacaga 420
Db 420 agagaataatgaatgactctgatgaatgacgatattattgaactacaagtgtcacaga 479
Qy 421 ctagaacttaacggaacagcttagagacag 450
Db 480 ctagaacttaacggaacagcttagagacag 509

RESULT 9
US-09-922-4808/c
; Sequence 4808, Application US/09359922
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4808
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-4808

Query Match 97.6%; Score 439; DB 51; Length 590;
Best Local Similarity 99.8%; Pred. No. 2.7e-79;
Matches 450; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 ctgctccagcgtgacgcccagccatggcgacgagagcttgaggcgctgaggagacag 60
Db 538 CTGCTCCAGCGCTGACGCGGAGCCATGGCGGACGAGGAGCTTGAGGCGCTGAGGAGACAG 479
Qy 61 aggcctggcagctcagcgcgaacacacgggacatcctgctgacgcccacaggaagca 120
Db 478 AGGCTGGCGAGCTGACGCCCAACACCGGGATCTGTGTGATGCGGCCCAACAGCA 419
Qy 121 aagcacaggaagcagaataagaaacagatctttagcccaagttctggatcagtcggcc 180
Db 418 AAGCACAGGGAAGCAGAAATGAGAAACAGATCTTGTAGCCCAAGTTCTGGATCAGTCGGCC 359
Qy 181 cgggcccaggttaagtaacttagcacttgaagcctgaaaaactaaagcagtagagaat 240
Qy 181 cgggcccaggttaagtaacttagcacttgaagcctgaaaaactaaagcagtagagaat 240
Db 358 CGGGCCAGGTTAAGTAACCTTAGCTTGTAAAGCCTGAAAAAAGCTTAAAGCAGTAGAGAA 299
Qy 241 tacctatagatggcagatgacaa-ctaagtgaagaggtatcagaaaggttt 299
Db 298 TACCTTATACAGATGCAAGATATGGACACACCTTAAGTGAAGGTATCAGAACAGGTTT 239
Qy 300 aatagaaatccttaaaaaagtaagcccaacaaacagaaagacaaacagtgaaattcaa 359
Db 238 AATAGAAATCCTTAAAAAAGTAAGCCCAACAAACAGAAAAAGACACACAGTGAATTCAA 179
Qy 360 cagaagaagaatgaatgactctgatgaatgacgattattgaactacaagtgtcacag 419
Db 178 CAGAAGAAAAGTAAATGGACTCTGATGAAGATGACGATTATTGAACACTACAAGTGTCTCACAG 119
Qy 420 actagaacttaacggaacagcttagagacag 450
Db 118 ACTAGAATTAACGGACAAAGTCTAGGACAG 88

RESULT 10
US-09-359-922-4808/c
; Sequence 4808, Application US/09359922A
; GENERAL INFORMATION:
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Liu, Jin
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922A
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: US 09/034,341
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4808
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-4808

Query Match 97.6%; Score 439; DB 51; Length 590;
Best Local Similarity 99.8%; Pred. No. 2.7e-79;
Matches 450; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 ctgctccagcgtgacgcccagccatggcgacgagagcttgaggcgctgaggagacag 60
Db 538 CTGCTCCAGCGCTGACGCGGAGCCATGGCGGACGAGGAGCTTGAGGCGCTGAGGAGACAG 479
Qy 61 aggcctggcagctcagcgcgaacacacgggacatcctgctgacgcccacaggaagca 120
Db 478 AGGCTGGCGAGCTGACGCCCAACACCGGGATCTGTGTGATGCGGCCCAACAGCA 419
Qy 121 aagcacaggaagcagaataagaaacagatctttagcccaagttctggatcagtcggcc 180
Db 418 AAGCACAGGGAAGCAGAAATGAGAAACAGATCTTGTAGCCCAAGTTCTGGATCAGTCGGCC 359
Qy 181 cgggcccaggttaagtaacttagcacttgaagcctgaaaaactaaagcagtagagaat 240
Db 358 CGGGCCAGGTTAAGTAACCTTAGCTTGTAAAGCCTGAAAAAAGCTTAAAGCAGTAGAGAA 299
Qy 241 tacctatagatggcagatgacaa-ctaagtgaagaggtatcagaaaggttt 299
Db 298 TACCTTATACAGATGCAAGATATGGACACACCTTAAGTGAAGGTATCAGAACAGGTTT 239
Qy 300 aatagaaatccttaaaaaagtaagcccaacaaacagaaagacaaacagtgaaattcaa 359
Db 238 AATAGAAATCCTTAAAAAAGTAAGCCCAACAAACAGAAAAAGACACACAGTGAATTCAA 179
Qy 360 cagaagaagaatgaatgactctgatgaatgacgattattgaactacaagtgtcacag 419
Db 178 CAGAAGAAAAGTAAATGGACTCTGATGAAGATGACGATTATTGAACACTACAAGTGTCTCACAG 119
Qy 420 actagaacttaacggaacagcttagagacag 450
Db 118 ACTAGAATTAACGGACAAAGTCTAGGACAG 88

RESULT 11
US-09-431-517-13611
; Sequence 13611, Application US/09431517
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-751CON1
; CURRENT APPLICATION NUMBER: US/09/431,517
; CURRENT FILING DATE: 1999-11-01
; EARLIER APPLICATION NUMBER: US 09/170,294
; EARLIER FILING DATE: 1998-10-13

```
; NUMBER OF SEQ ID NOS: 31760
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13611
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(421)
; OTHER INFORMATION: n = A,T,C or G
US-09-431-517-13611

Query Match      85.4%; Score 384.2; DB 55; Length 421;
Best Local Similarity 96.6%; Pred. No. 2.8e-68;
Matches 400; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

Qy 28 gcgagcagagagcttgaggcgctgagagacagagagctgcccagctgcagggccaaacac 87
Db 8 gtggcgagagagcttgaggcgctgagagacagagagctgcccagctgcagggccaaacac 67
Qy 88 ggggatcctggtgatgctggcgcccaacaggaagcaaacgacaggggaagcagaatgagaac 147
Db 68 ggggatcctggtgatgctggcgcccaacaggaagcaaacgacaggggaagcagaatgagaac 127
Qy 148 agtatcttagcccaagtcttgatcagtcggcccgccgaggttaagttaacttagcactt 207
Db 128 agtatcttagcccaagtcttgatcagtcggcccgccgaggttaagttaacttagcactt 187
Qy 208 gtaagcctgaaaaactaaagcagtagaattacaccttatacagatggcaagatatgga 267
Db 188 gtaagcctgaaaaactaaagcagtagaattacaccttatacagatggcaagatatgga 247
Qy 268 caactaagtgaaggtatcagaacaaggtttaataagaaatccttaaaaaagtaagccaa 327
Db 248 caactaagtgaaggtatcagaacaaggtttaataagaaatccttaaaaaagtaagccaa 307
Qy 328 caaacagaaaaagacaacaacagtgaaattcaacagagaagaagtaagtctgatgaa 387
Db 308 caaacagaaaaagacaacaacagtgaaattcaacagagaagaagtaagtctgatgaa 367
Qy 388 gatcacgattattgaactacaagtgctcacagactagaacttaacgggaacaagt 441
Db 368 gatcacgattattgaactnccagggttcgcgnaacta-aacttancggaccaagt 420

RESULT 13
US-09-431-517-13610
; Sequence 13610, Application US/09431517
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-751CON1
; CURRENT APPLICATION NUMBER: US/09/431,517
; CURRENT FILING DATE: 1999-11-01
; EARLIER APPLICATION NUMBER: US 09/170,294
; EARLIER FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 31760
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13610
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(420)
; OTHER INFORMATION: n = A,T,C or G
US-09-431-517-13610

Query Match      82.7%; Score 372; DB 55; Length 420;
Best Local Similarity 97.1%; Pred. No. 8e-66;
Matches 397; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

Qy 28 gcgagcagagagagcttgaggcgctgagagacagagagctgcccagctgcagggccaaacac 87
Db 8 gtggcgagagagcttgaggcgctgagagacagagagctgcccagctgcagggccaaacac 67
Qy 88 ggggatcctggtgatgctggcgcccaacaggaagcaaacgacaggggaagcagaatgagaac 147
Db 68 ggggatcctggtgatgctggcgcccaacaggaagcaaacgacaggggaagcagaatgagaac 127
Qy 148 agtatcttagcccaagtcttgatcagtcggcccgccgaggttaagttaacttagcactt 207
Db 128 agtatcttagcccaagtcttgatcagtcggcccgccgaggttaagttaacttagcactt 187
Qy 208 gtaagcctgaaaaactaaagcagtagaattacaccttatacagatggcaagatatgga 267
Db 188 gtaagcctgaaaaactaaagcagtagaattacaccttatacagatggcaagatatgga 247
Qy 268 caactaagtgaaggtatcagaacaaggtttaataagaaatccttaaaaaagtaagccaa 327
Db 248 caactaagtgaaggtatcagaacaaggtttaataagaaatccttaaaaaagtaagccaa 307
Qy 328 caaacagaaaaagacaacaacagtgaaattcaacagagaagaagtaagtctgatgaa 387
Db 308 caaacagaaaaagacaacaacagtgaaattcaacagagaagaagtaagtctgatgaa 367
Qy 388 gatcacgattattgaactacaagtgctcacagactagaacttaacgggaacaagt 441
Db 368 gatcacgattattgaactnccagggttcgcgnaacta-aacttancggaccaagt 420

RESULT 12
US-09-431-517-14582
; Sequence 14582, Application US/09431517
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-751CON1
; CURRENT APPLICATION NUMBER: US/09/431,517
; CURRENT FILING DATE: 1999-11-01
; EARLIER APPLICATION NUMBER: US 09/170,294
; EARLIER FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 31760
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14582
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(421)
; OTHER INFORMATION: n = A,T,C or G
US-09-431-517-14582

Query Match      85.4%; Score 384.2; DB 55; Length 421;
Best Local Similarity 96.6%; Pred. No. 2.8e-68;
```

```
Db 188 gtaaacctgaaaaactaaagcagtagagaattaccttatcacagatggcgaagatatga 247
QY 268 caactaagtgaaggtatcagacaaggtttaatagaaatccttaaaaaagtaagccaa 327
Db 248 caactaagtgaaggtatcagacaaggtttaatagaaatccttaaaaaagtaagccaa 307
QY 328 caaacgaaaaagacaaacacagtcgaaattcaacagagaagaaagttaaggac-tctgatga 386
Db 308 caaacgaaaaagacaaacacagtcgaaattcaacagagaagaaagttaaggcctttgntga 367
QY 387 agatgacgattattgaactaacagtcgtcacagactagaaacttaacgga 435
Db 368 anatacgattnttgactacaagtgtc-cagactanaacttaacgga 415
```

```
RESULT 14
US-09-362-510-58828
; Sequence 58828, Application US/09362510
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-759CON1
; CURRENT APPLICATION NUMBER: US/09/362,510
; CURRENT FILING DATE: 1999-07-27
; EARLIER APPLICATION NUMBER: US 09/221,820
; NUMBER OF SEQ ID NOS: 62165
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 58828
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(476)
; OTHER INFORMATION: n = A,T,C or G
US-09-362-510-58828
```

```
Query Match 80.4%; Score 362; DB 52; Length 476;
Best Local Similarity 98.6%; Pred. No. 8.4e-64;
Matches 365; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 tgcctcagcgtgaacgcgcgagccatggcgacagagagcttgaggcgctgagagacaga 61
Db 106 ttctccagctttgacgcgcgagccatggcgacagagagcttgaggcgctgagagacaga 165
QY 62 ggcctggccgagctgcagggccaaacacacggggtcctggtgatgcggccacaggaagcaa 121
Db 166 ggcctggccgagctgcagggccaaacacacggggtcctggtgatgcggccacaggaagcaa 225
QY 122 agcacagggaagcagaataagaaacagtcattcttagcccaagttctggatcagtcggccc 181
Db 226 agcacagggaagcagaataagaaacagtcattcttagcccaagttctggatcagtcggccc 285
QY 182 gggccaggttaagttaactagcacttgaagcctttaaagcctgaaaaaactaaagcagtagagaatt 241
Db 286 gggccaggttaagttaactagcacttgaagcctttaaagcctgaaaaaactaaagcagtagagaatt 345
QY 242 acctatcacagatggcagaataggaactatggaactgaagtgagaggttatcagaacaaggtttaa 301
Db 346 acctatcacagatggcagaataggaactatggaactgaagtgagaggttatcagaacaaggtttaa 405
QY 302 tagaaatccttaaaaaagtaagccaaacacagagaagaaacacacagtcgaaattcaaca 361
Db 406 tagaaatccttaaaaaagtaagccaaacacagagaagaaacacacagtcgaaattcaaca 465
QY 362 gaagaaaagt 371
Db 466 gaagaaaagt 475
```

```
RESULT 15
US-60-172-373-25470
; Sequence 25470, Application US/60172373
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE REFERENCE: GX-0006 P
; CURRENT APPLICATION NUMBER: US/60/172,373
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 25,772
; SOFTWARE: PERL Program
; SEQ ID NO 25470
; LENGTH: 626
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 235725.4
US-60-172-373-25470

Query Match 80.2%; Score 361; DB 92; Length 626;
Best Local Similarity 100.0%; Pred. No. 1.4e-63;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 ggatcctggtgatgcggcccaacagcaagcaagcagggggaagcagagaaatgagaaaaacag 149
Db 232 ggatcctggtgatgcggcccaacagcaagcaagcagggggaagcagagaaatgagaaaaacag 291
QY 150 tatcttagcccaagttctggtatcagtcgcccggccaggttaagttaacttagcacttgt 209
Db 292 tatcttagcccaagttctggtatcagtcgcccggccaggttaagttaacttagcacttgt 351
QY 210 aaagcctgaaaaaactaaagcagtagagaattaccttatcacagatggcgaagatatggaca 269
Db 352 aaagcctgaaaaaactaaagcagtagagaattaccttatcacagatggcgaagatatggaca 411
QY 270 actaagtgaagsggttatcagaacaaggttttaataagaaatccttaaaaaagtaagccaaca 329
Db 412 actaagtgaagsggttatcagaacaaggttttaataagaaatccttaaaaaagtaagccaaca 471
QY 330 aacagaaaaagacaacaacacagtgaaattcaacagagaagaaagtaatggactctgatgaaga 389
Db 472 aacagaaaaagacaacaacacagtgaaattcaacagagaagaaagtaatggactctgatgaaga 531
QY 390 tgacgattattgaactacaagtgctcacagactagaacttaacgcggaacaagtcctaggaca 449
Db 532 tgacgattattgaactacaagtgctcacagactagaacttaacgcggaacaagtcctaggaca 591
QY 450 g 450
Db 592 g 592
```

Search completed: May 24, 2000, 01:01:46
Job time: 14278 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:27:38 ; Search time 6115.43 Seconds
(without alignments)
-186.591 Million cell updates/sec

Title: US-09-215-435-117
Perfect score: 1173
Sequence: 1 gagctgcttatgggacacg.....tccatctcaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_cm.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_vi.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_sy.*
- 29: em_un.*
- 30: em_vi.*
- 31: gb_htg1.*
- 32: gb_htg2.*
- 33: gb_in1.*
- 34: gb_in2.*
- 35: gb_ba1.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_pl3.*
- 51: gb_pr5.*
- 52: gb_htg8.*
- 53: gb_htg9.*
- 54: gb_htg10.*
- 55: gb_htg11.*
- 56: gb_htg12.*
- 57: gb_htg13.*
- 58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
c 1	137	11.7	67883	8	SCE8229
c 2	106.6	9.1	149376	41	AC009258
c 3	97.8	8.3	95628	55	AC008254
c 4	97.6	8.3	922	13	CNS01G20
c 5	93.8	8.0	123579	43	AC017792
c 6	86.2	7.3	720	8	CNS01A07
c 7	76.4	6.5	18700	8	SPAC694
c 8	67	5.7	24775	34	CEC37H6
c 9	66.8	5.7	83689	7	AB017067
c 10	53.6	4.6	80367	8	AFAC012329
c 11	53.6	4.6	129757	8	ATF2K15
c 12	52.6	4.5	194613	41	AC009741
c 13	47.8	4.1	12460	2	AF001311
c 14	47	4.0	39339	34	CEK08H10
c 15	44.4	3.8	720	8	CNS01B5Z
c 16	42	3.6	48849	55	AC023203
c 17	40.8	3.5	55482	54	AC022663
c 18	40.4	3.4	41421	11	AC004221
c 19	40.2	3.4	169027	33	AC002370
c 20	40	3.4	1958	12	AB017545
c 21	39.2	3.3	142299	32	AP001082
c 22	39.2	3.3	146467	41	AC009438
c 23	39	3.3	1246	1	TTHPFK1A
c 24	38.8	3.3	2156	1	SCU19249
c 25	38.8	3.3	38084	1	SCE68
c 26	38.8	3.3	139779	33	AL136528
c 27	38.6	3.3	12347	2	AE002051
c 28	38.4	3.3	1247	11	AF045186
c 29	38.4	3.3	1908	9	AB017546
c 30	38.4	3.3	149932	33	AL139423
c 31	38	3.2	1191	35	AF160913
c 32	38	3.2	1215	35	AF132155
c 33	38	3.2	3131	12	AF155372
c 34	38	3.2	3168	12	AF155373
c 35	38	3.2	63855	43	AC016794
c 36	38	3.2	89787	44	AC020266
c 37	38	3.2	115847	42	AC011696
c 38	38	3.2	162471	43	AC009908
c 39	38	3.2	165202	44	AC016877
c 40	38	3.2	172906	44	AF216671
c 41	38	3.2	194859	41	AC007473
c 42	37.8	3.2	1136	8	AF112964
c 43	37.8	3.2	127339	12	AC000096
c 44	37.8	3.2	166820	33	HSB4425A6
c 45	37.8	3.2	168302	54	AC008271

ALIGNMENTS

```

RESULT 1
LOCUS SCE8229 67883 bp DNA PLN 01-AUG-1997
DEFINITION Saccharomyces cerevisiae chromosome V cosmid 8229, 9115, 9132, 9811, and lambda clones 7990 and 6134.
ACCESSION U18917 L10718 U00092
VERSION U18917.1 GI:603377
KEYWORDS SPT15 (TATA-box binding protein tfIID); OXAL (involved in cytochrome oxidase biogenesis); RAD4 (nucleotide excision repair protein); BEM2 (bud-emergence protein); PET122 (translational activator of cytochrome c oxidase subunit III); MAG1 (3-methyladenine DNA glycosylase); SPT2 (non-specific DNA binding protein); UBP3 (ubiquitin-specific protease); UBP5 (ubiquitin-specific protease); tRNA-Val; tRNA-Arg; tRNA-Glu; tRNA-Ile.

SOURCE
  baker's yeast.
  Saccharomyces cerevisiae
  Eukaryota; Fungi; Ascomycota; Saccharomycetales;
  Saccharomycetaceae; Saccharomyces.
REFERENCE
  1 (bases 1 to 67883)
  Dietrich,F.S., Mulligan,J., Hennesy,K., Allen,E., Araujo,R.,
  Aviles,E., Berno,A., Brennan,T., Carpenter,J., Chen,E.,
  Cherry,J.M., Chung,E., Duncan,M., Guzman,E., Hartzell,G.,
  Hunnicke-Smith,S., Hyman,R., Kayser,A., Komp.C., Lashkari,D.,
  Lew,H., Lin,D., Mosedale,D., Nakahara,K., Namath,A., Norgren,R.,
  Oefner,P., Oh,C., Petel,F.X., Roberts,D., Sehl,P., Schraumm,S.,
  Shogren,T., Smith,V., Taylor,P., Wei,Y., Yelon,M., Botstein,D. and
  Davis,R.W.
  The complete sequence of Saccharomyces cerevisiae chromosome V
  Unpublished (1994)
  2 (bases 1 to 67883)
  Dietrich,F.S.
  The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9811, and
  lambda clones 7990 and 6134
  Unpublished (1994)
  3 (bases 1 to 67883)
  Dietrich,F.S.
  Direct Submission
  Submitted (20-DEC-1994)
  4. (bases 1 to 67883)
  Jia,Y. and Cherry,J.M.
  Direct Submission
  Submitted (20-MAY-1997) Department of Genetics, Stanford
  University, Saccharomyces Genome Database, Stanford, CA 94305-5120,
  USA
COMMENT
  On Jul 19, 1996 this sequence version replaced gi:172801.

Sequenced by: Stanford DNA Sequence & Technology Center
              855 California Avenue
              Palo Alto, CA 94304, USA

Curated by: Saccharomyces Genome Database
             URL: http://genome-www.stanford.edu/
             e-mail: yeast-curator@genome.stanford.edu

Neighboring Sequence:
The 5' end of this sequence overlaps with GenBank Accession Number
U18916.
FEATURES
  source
    1. .67883
    /location/Qualifiers
    /organism="Saccharomyces cerevisiae"
    /strain="S288C (AB972)"
    /db_xref="taxon:4932"
    /chromosome="V"
    /map="5K"
    <1. .23020
    /organism="Saccharomyces cerevisiae"
    /db_xref="taxon:4932"
    /clone="cosmid 9115"
    <1. .400

  source
    9549. .11219
    /gene="YER140W"
    9549. .11219

  source
    complement(1192. .1265)
    /note="trna-ile; codon recognized: ATT"
    /product="trna-glu"
    complement(1405. .1718)
    /note="delta element"
    1470. .1567
    /note="sequence from one strand only"
    1567. .42183
    /organism="Saccharomyces cerevisiae"
    /db_xref="taxon:4932"
    /clone="cosmid 8229"
    complement(1741. .7009)
    /gene="YER138C"
    complement(join(1741. .5704,5706. .7009))
    /gene="YER138C"
    /note="transposon Ty putative with frame shift at
    5704-5706; Yer138cp"
    /codon_start=1
    /evidence=not_experimental
    /protein_id="AAB64665.1"
    /db_xref="GI:603378"
    /translation="MESQQLSQHSPISHGSCASVTSKEVHTNODPLDVSASKTECE
    KASTKANSQTTTPASSAVPENPHASPPASVPPQNGPIPPQCMNTQNGNSGWS
    FYGHPMTPYQPMSPMYPFPQSPQFPYSSVPLSTPSPESGNTFYDSSADS
    DWTSTKTVRRPMLTSPNDPNWVTVYKFLQNSLGGIITPVNGKPVQITDDEL
    FLYTFOIFAPSQFLPTWVDIMKILSKSIEKMQSDQANDQIVTLANLQ
    YNGSTPADAFETKVTNIIDRLNNGIHNNKACQLMRGLSGYKFLYRTRHLNM
    TVAEFLDILHAIYEEQGSRSKNYRNRLSDEKNSRSYNTTKPKVIARNPKNTN
    SKSTARAHNVSTSNNSPSTNDLSKSTETPEIQLNNKHDHLGQELTESYVNTNHS
    DDELPGHLLDSGASRTLSRAHHHSASSNPDI NVDAQRDGTVPINAIQDFHQD
    NTKYKLVHTNPATYDILLSNELAAVDITACFTKNVLRSDGVVLAPIVYGFYV
    SKVLLPNSISVPTNNVHTSESTKRYKPIHRMLAHANAQTI RYSLKNTITFNE
    SDVWSSAIDQCPCLLGSKTKHRIKGRKLYQNSYEPFOYLHTDIFGVHNLPS
    APSFISDTETRFWYPLHDREDSILDVFTILAF IKNOQASVLVIQMDRGE
    YNRTLHFLKNGITPYTTTADSRAGVAERLNRLLDCRTQLOQSGSLPNLWFS
    AIEFTIVRNSLASPKSAQHAGLADLSTLLPGQPVYVNDHNPNSKIHPRGI
    PGYALHPSRNSYGIYILSLKTKVDTTNYILOKESRLDOFNDALTFDDELNRLT
    ASYQSIASNEIQDILNIESDHFQSDIELHPQPRNVLSKAVSPDTSPPSHTE
    DSKRYSKTNIRAPREVDPNIESNLSPKSRSTPQISNIESTGGMKLNVLPLAP
    MSQNTHESSHASKSKDRHSDSYSENETHNTVPISSTGTNNKTVPOIQDQTEKR
    ITHRSPDASPPENNSHNVPKPTTVTSEQNTESIADLPDLPDPSPESPFPD
    PKELPPINSHOTSLSLIGSDSNAYTTINSKRSLEDNETEIKVSRDTWTKNRSL
    EPRSKRIHQAAVKSKIPRTITLYDEAITYNKDIKEKIKIICAYHKEYNQLL
    KMTWDTDEYDRKEIDPKRVINSMFIFKKRQDTHKARFARGVDIQHPDYDSMQS
    NTVHYALMTSLALDNNYITQDISAYLYADIKBELYIRPPHGLMMDKLIRLK
    KSLYGLKSGANNYETIKSLIKONGEIOYDILGLEIKVORGKYMKGMSLTKIP
    NANKLIITLKKQYDTKILNGESONEIOYDILGLEIKVORGKYMKGMSLTKIP
    KUNVPLNPKRKLAPGQPLVIDODELEIDEDEYKVKHEMQKILGASVGYKFRF
    DILYYINTLAQHILFSPQVLDWMTYELIQFMDTRDQLIHNNKPTPEDNKLVAISD
    ASYQNPQYKQSIGNIYLLNGKVIQGGTKASLTCTCTEAEIHAISESVPLLNLSY
    LIQELNKKPIIKGLLTDSDSTSIISKSTNEERFRNFRFTGTAMRLRDEVSNGNLVYVY
    IETKNIADVMTKPLPKTKLTLTKNWIH"
    complement(8547. .9227)
    /gene="YER139C"
    complement(8547. .9227)
    /note="contains motifs typical of ATP/GTP-binding sites"
    /gene="YER139C"
    /codon_start=1
    /evidence=not_experimental
    /product="Yer139cp"
    /protein_id="AAB64666.1"
    /db_xref="GI:603379"
    /translation="MATIEDIKETALIPQKHQRLQSMHEAVITLIEIGLLCDSECKD
    EKTLYLGRFLTPDMYQDLVDRNLNKRGPYLCGKSPERIRDPMDMTTKKFLLEN
    NPAYLSHYCSKHFRCQYQVQSLDEALFARTGVHLFEDPEQKHIDIDFKVTLFEE
    LLREKASEEDIKLSLGLKGLNPDPSGTTEKDDTELEDLDSKLQAIKIVENDNPSI
    LGDFTRED"
    9549. .11219
    /gene="YER140W"
    9549. .11219
  
```

```

/gene="YER140w"
/codon_start=1
/evidence-not_experimental
/product="Yer140wp"
/protein_id="AAB64667.1"
/db_xref="GI:603380"
/translation="MQHKDTAVAKDTAKKRLRRNSAPSIAHIIISRLDKWFLWNTI
DRNIIVEQDESSAAKEEHEDDYLEQLLNIRIPMFLEKFMFLALITSLDCLFY
FTVPLRLIKYVQKPSYRHOYRLOQRSGHKNKIPFRYRITSREYKERCMIRIIVIS
SILLKLDISKLYHRKROSTMKLYMLFVLEWADKMLASLGSLTVMLSKNSERI
LLHKCLLVNSLIYVTHIGVIVYQALISUNIAVNSYNALLTLLLSQFAEIKSVLK
KFKDEGFTQITADVVERFKLITLLSITGLRNLQSSSSLSNITSINFWSPRSLTSLVI
NILCGPVSVYSGEVDWAKHAYITKFNRIQYIDKFYIIYIKDYSTRHKLDELRL
GLPLPAFVFLFVWVRPTLFKSSSEPVLPFLPMGASVFLALLALHSDYKNENRTESV
WSKRIEQRFRDQAFNTVTEEVVPGLLSGMGKVDVSTRIALHSDYKNENRTESV
SPMKRKTILTACTPESLNDIRQRDKNPRSLNENYARKVYKWSKRIW"
11443..12903
/gene="COX15"
/cds_start=1
11443..12903
/gene="COX15"
/notes="YER141w"
/codon_start=1
/evidence-experimental
/product="Cox15p: Cytochrom oxidase assembly factor"
/protein_id="AAB64668.1"
/db_xref="GI:603381"
/translation="MLFRNTEVGRQAAKLLTRSSRLAWOSIGASRNISTIROIRKT
OLYNFKTYSIRPFLSSPVFKPHVASESNPIESRLKTSKNVAYWLLGTSLVFGIVV
LGLTRTESGLSITWPKPVTGLPPMNQKEEIEFKYKESPEFLNLSHIDIDDFK
FIFFMWHLRGAIRGAVFILPAVFAVSKKTSGHVKNRKLGLGLLGLQGVGWMW
VKSGLDQDLARKSPKTSQVYRLTHLTAFFLYMGLWLTGLEILRECKWIKNPVQA
ISLFLKLDNPAIGPMRKISALLAVSFLTAMSGMVAGLDAGWVYNTWPKWGERWPS
SRELDENCFRRDKDLWRNLENPVTQLVHRTCAVVAFTSVLAAHMYAIKKAV
IPRNATSLHVMGVYTLQATLIGLITLILYPLISIASIHQAGALALLTSLVFPASQLR
KPRAPKRNVIITPHSKSVTSGKILSEASKLAKPL"
complement(13130..14020)
/gene="MAG1"
/db_xref="SGD:S0000944"
complement(13130..14020)
/gene="MAG1"
/notes="YER142c"
/codon_start=1
/db_xref="SGD:S0000944"
/product="Mag1p: 3-methyladenine DNA glycosylase"
/protein_id="AAB64669.1"
/db_xref="GI:603382"
/translation="MKLREYDELIRADAVEIAKEIGSRPLEVALPEKVIARHEKPF
NMACEHILEKDPISLFLPKNNEFTLYKETQVPNTLEDYFIRLASTILSQOISQAAE
SIKARVYSLGGAPPDYKILFEDFKPAKCAEIAKGLSKRMIIYLESIAVYFTEKYK
DIEKLFQKDNDEVTESLVTVNKGIPWSAKMFLISGLKRMDFVAFEDLGIARGFSK
YLSDKPPELEKELRERKVKKSKIKHKYNNKIYDDDDIMEKCSSETSPFSVPMFILW
RLASTNTDAMKAEENFVS"
14303..15589
/gene="YER143w"
14303..15589
/gene="YER143w"
/codon_start=1

```

Query Match 11.78; Score 137; DB 8; Length 67883;

Best Local Similarity 49.19; Pred. No. 1.4e-23;

Matches 488; Conservative 0; Mismatches 460; Indels 45; Gaps 3;

Qy 145 cgaatcggagcaacatggcaccttcacatgcgacgaggaactggcactgctgctt 204

Db 42280 CAAATATGACGCATTCGGGATCTTTCACGCGGACGAGCTGTTGGCGGTATACATGTTA 42221

Qy 205 cgcctctcgcgagtagccggatgcagagatgtgcccggaccggggtcccgaaaaactc 264

Db 42220 AGGCTACTACCCGATTTAAAGATGCTAAGTGTGTAGTCAAGAAATCCCAAGACTGG 42161

Qy 265 gcttctgtgacatgtgtgtgagcgtgggggagtcagcaccctcgagagacccgatat 324

Db 42160 GAGGCTAGTACATCTTCTAGTGTAGGAGCTCAGTATGACGGCGTGAATTTTTCGAC 42101

```

Qy 325 gaccatcaccagaggtcttccacagagaccatgagctccctgtccccctgggagcgctgg 384
Db 42100 CACCACACCGTGGATTTTCGAAACATTTCAACGA-----GAAATAC 42059
Qy 385 cagaccaagctgagcagtcgaggggactcatctatctgcacttcggtcggcacaaactgctggcc 444
Db 42058 AAGACTAAGCTCTCAAGTCTGGTCTGATCTTCAACACTACGTCGTCGACATTATCAA 41999
Qy 445 cagttgctgggacactagtgaagagagacagcatggtgggacccctctctatgacaagaatgat 504
Db 41998 ACAATTTTGACAAACAGGTTAGCAGCTCTGATTAGATCTGCTCTATGACAAGATATAC 41939
Qy 505 gagaactttgtggagaggtggtgacgtgctggtgacaaatggatc----- 547
Db 41938 AACAATTTGTGAGGACACTGGATGCAACACGCAACGGTATTAGCAAGTACACGATTCCT 41879
Qy 547 -----tcccagtgggcagagggggagcctcgatgcactgacacactaccctgagtgc 600
Db 41878 AAGGATTCCAACTTGAACACCAATTTTCAGGGATAACGCCATCAGCATTCCTCCGGGATTAT 41819
Qy 601 cgattgctgcacttaactcactggaacccaccccgacacacactgagggaggttc 660
Db 41818 TCAGGTATGAATCCTTAACCTGACCAAGACACATTCCTCGATGAGTCTTTGACAGATGCTTT 41759
Qy 661 aagcgtgcaatggatcgtgtcaagagaggtttctgcagagattagattttaccacac 720
Db 41758 GCTCGCGCAAGTAGATTATTCGTTGGGGTCTTTGTCACATTTGTAAGAGGTACGGTCAA 41699
Qy 721 agctggtgccagcccgcccttggtagaagagcccttgcacagcattccagtggtgac 780
Db 41698 TCGTGGTTGCCAGCAAGGCCCTTGTGGCGCAAGCCATTGATGAAAGAATGGATGTTGAT 41639
Qy 781 ccaagtggagagattgtggaactggcgaagagtcagtcctcctggaagagagacatctctac 840
Db 41638 AAGAGTGGCAAAATCATTTGCTTGGCCCCAA---TTCTGTCCATGGAAGGACATCTATAC 41582
Qy 841 cactggaactctggctgctccctccagtgccactcttctgttatctacacagaccag 900
Db 41581 GAGCTGGAAAGAGAGAGAACAATTTGMAAAGCAAAATCGAGTTTGTCTTCACAGACTCT 41522
Qy 901 gctggcagtgccgaatacacagtgtgtgccaaaggagccccactcattccaaagccgctg 960
Db 41521 TCTGTTGCTGGAGGGTATCTACTGTACCAATCAACTCCACCTCTCTCCAGTTTAGGAGA 41462
Qy 961 cccctgccagagcactggcgggtcttctggagcagagcccttgaccaggtcagtggtgac 1020
Db 41461 GGTTCGCGGAGCCATTAAAGAGGCTTCAGAGACGAAGAGCTGAGCACCAGAGCGGTGTG 41402
Qy 1021 cctgctgcactctcgtccatgcaagcggcttcattgctgcgtaccgcaccccgagaggtt 1080
Db 41401 CCAGATGATATCTTCATCTCAGCAGAGTTTCATTTGAGGTGCCAAGTCCCAAGAGGCT 41342
Qy 1081 gccttgagcatggccgtgcccactctggccacg 1113
Db 41341 GTATATGAGCTGGCTAAGATGCTTTTGGCCTAG 41309

```

RESULT 2

AC009258

LOCUS

DEFINITION

AC009258

AC009258

VERSION

AC009258.2

KEYWORDS

HTG; HTGS_PHASE1.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 149376)

AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE	Sequencing of Drosophila melanogaster
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 149376)
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE	Direct Submission
JOURNAL	Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT	On Aug 17, 1999 this sequence version replaced gi:5713326. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (http://www.fruitfly.org/sequence/) or send email to bdgs@fruitfly.berkeley.edu . All contigs in this submission meet the following cutoffs: length >= 200 bases. * NOTE: This is a 'working draft' sequence. It currently * consists of 113 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 1137: contig of 1137 bp in length * 1138 1217: gap of unknown length * 1218 2420: contig of 1203 bp in length * 2421 2500: gap of unknown length * 2501 3119: contig of 619 bp in length * 3120 3199: gap of unknown length * 3200 3941: contig of 742 bp in length * 3942 4021: gap of unknown length * 4022 4808: contig of 787 bp in length * 4809 4888: gap of unknown length * 4889 5813: contig of 925 bp in length * 5814 5893: gap of unknown length * 5894 6584: contig of 691 bp in length * 6585 6664: gap of unknown length * 6665 7461: contig of 797 bp in length * 7462 7542: gap of unknown length * 7543 8084: contig of 542 bp in length * 8084 8164: gap of unknown length * 8164 8834: contig of 670 bp in length * 8834 9591: gap of unknown length * 9591 9671: contig of 878 bp in length * 9672 10275: gap of unknown length * 10276 10355: gap of unknown length * 10356 11056: contig of 701 bp in length * 11057 11136: gap of unknown length * 11137 11836: contig of 700 bp in length * 11837 11916: gap of unknown length * 11917 12415: contig of 499 bp in length * 12416 12495: gap of unknown length * 12496 13429: contig of 934 bp in length * 13430 13509: gap of unknown length * 13510 14534: contig of 1025 bp in length * 14535 14614: gap of unknown length * 14615 15411: contig of 797 bp in length * * 15412 15491: gap of unknown length * 15492 16096: contig of 605 bp in length * 16097 16176: gap of unknown length * 16177 16925: contig of 748 bp in length * 16926 17004: gap of unknown length * 17005 18033: contig of 1029 bp in length * 18034 18113: gap of unknown length * 18114 18637: contig of 524 bp in length * 18638 18717: gap of unknown length * 18718 19318: contig of 601 bp in length * 19319 19398: gap of unknown length * 19399 20124: contig of 726 bp in length * 20125 20205: gap of unknown length * 20206 20799: contig of 595 bp in length * 20800 21699: contig of 820 bp in length * 21700 21779: gap of unknown length * 21780 22694: contig of 914 bp in length * 22695 22774: gap of unknown length * 22775 24639: contig of 1865 bp in length * 24640 24719: gap of unknown length * 24720 25591: contig of 873 bp in length * 25592 25671: gap of unknown length * 25672 26769: contig of 1098 bp in length * 26770 26849: gap of unknown length * 26850 28057: contig of 1208 bp in length * 28058 28137: gap of unknown length * 28138 28612: contig of 1475 bp in length * 28613 28692: gap of unknown length * 28693 30564: contig of 872 bp in length * 30565 30644: gap of unknown length * 30645 31815: contig of 1171 bp in length * 31816 31895: gap of unknown length * 31896 32761: gap of unknown length * 32762 32841: contig of 1228 bp in length * 32842 34069: gap of unknown length * 34070 35148: contig of 1000 bp in length * 35149 35228: gap of unknown length * 35229 36191: contig of 963 bp in length * 36192 36271: gap of unknown length * 36272 37401: contig of 1130 bp in length * 37402 37481: gap of unknown length * 37482 38929: contig of 1448 bp in length * 38930 39796: gap of unknown length * 39797 39876: gap of unknown length * 39877 40805: contig of 929 bp in length * 40806 40885: gap of unknown length * 40886 41986: gap of unknown length * 41987 42065: gap of unknown length * 42066 44189: contig of 2124 bp in length * 44190 44269: gap of unknown length * 44270 46357: contig of 2088 bp in length * 46358 46437: gap of unknown length * 46438 48288: contig of 1850 bp in length * 48289 50437: gap of unknown length * 50438 50436: contig of 2069 bp in length * 50437 50516: gap of unknown length * 50517 51741: contig of 1225 bp in length * 51742 51821: gap of unknown length * 51822 53193: contig of 1371 bp in length * 53194 53272: gap of unknown length * 53273 55250: contig of 1978 bp in length * 55251 55330: gap of unknown length * 55331 56487: contig of 1157 bp in length * 56488 58624: gap of unknown length * 58625 58704: gap of unknown length * 58705 60336: contig of 1632 bp in length * 60337 62682: contig of 2266 bp in length * 62683 62762: gap of unknown length

* 62763	64401: contig of 1639 bp in length	64401: contig of 1639 bp in length
* 64402	64481: gap of unknown length	64481: gap of unknown length
* 64482	66687: contig of 2206 bp in length	66687: contig of 2206 bp in length
* 66688	66767: gap of unknown length	66767: gap of unknown length
* 66768	69470: contig of 2703 bp in length	69470: contig of 2703 bp in length
* 69471	69550: gap of unknown length	69550: gap of unknown length
* 69551	72548: contig of 2998 bp in length	72548: contig of 2998 bp in length
* 72549	72628: gap of unknown length	72628: gap of unknown length
* 72629	74611: contig of 1983 bp in length	74611: contig of 1983 bp in length
* 74612	74691: gap of unknown length	74691: gap of unknown length
* 74692	76799: contig of 2108 bp in length	76799: contig of 2108 bp in length
* 76800	76879: gap of unknown length	76879: gap of unknown length
* 76880	78964: contig of 2085 bp in length	78964: contig of 2085 bp in length
* 78965	79044: gap of unknown length	79044: gap of unknown length
* 79045	81321: contig of 2277 bp in length	81321: contig of 2277 bp in length
* 81322	81401: gap of unknown length	81401: gap of unknown length
* 81402	83832: contig of 2431 bp in length	83832: contig of 2431 bp in length
* 83833	83912: gap of unknown length	83912: gap of unknown length
* 83913	85967: contig of 2055 bp in length	85967: contig of 2055 bp in length
* 85968	86047: gap of unknown length	86047: gap of unknown length
* 86048	88677: contig of 2630 bp in length	88677: contig of 2630 bp in length
* 88678	88757: gap of unknown length	88757: gap of unknown length
* 88758	91009: contig of 2252 bp in length	91009: contig of 2252 bp in length
* 91010	91089: gap of unknown length	91089: gap of unknown length
* 91090	94632: contig of 3543 bp in length	94632: contig of 3543 bp in length
* 94633	94712: gap of unknown length	94712: gap of unknown length
* 94713	98030: contig of 3318 bp in length	98030: contig of 3318 bp in length
* 98031	98110: gap of unknown length	98110: gap of unknown length
* 98111	104063: contig of 5953 bp in length	104063: contig of 5953 bp in length
* 104064	104143: gap of unknown length	104143: gap of unknown length
* 104144	109813: contig of 5670 bp in length	109813: contig of 5670 bp in length
* 109814	109893: gap of unknown length	109893: gap of unknown length
* 109894	120039: contig of 10146 bp in length	120039: contig of 10146 bp in length
* 120040	120119: gap of unknown length	120119: gap of unknown length
* 120120	120773: contig of 654 bp in length	120773: contig of 654 bp in length
* 120774	120853: gap of unknown length	120853: gap of unknown length
* 120854	121499: contig of 646 bp in length	121499: contig of 646 bp in length
* 121500	121579: gap of unknown length	121579: gap of unknown length
* 121580	122201: contig of 622 bp in length	122201: contig of 622 bp in length
* 122202	122281: gap of unknown length	122281: gap of unknown length
* 122282	122995: contig of 714 bp in length	122995: contig of 714 bp in length
Query Match 9.1%; Score 106.6; DB 41; Length 149376;		
Best Local Similarity 50.6%; Pred. No. 3.9e-16;		
Matches 312; Conservative 0; Mismatches 299; Indels 6; Gaps 2;		
QY 479	tgggacccctctatgacaagatgtatgagaacttttggaggaggttgatgctgttgaca	538
Db 76128	TGCAAAACGGCGCTTATTCAGATCTATAGGAATTTTATCAACGAACCTGGACGCAATTGATA	76187
QY 539	atgggatctcccaatgggcagaggggagcctcgatgacactgaccactaccctgagtg	598
Db 76188	ATGGAGTGGCCATGTTCGAGGGCGTTGAGCGCTATATACAAAATCTCCATCTATCTG	76247
QY 599	cacgagtgtgctactaatcctcactctggaaccacccacacacactgagcaggggt	658
Db 76248	CCGCGATTGCCAACTGAATCCGCTCTGCGAGGAACCTGGCTGAGACATCGAAGACCGAT	76307
QY 659	tcaagcgtgcaatggatctggttcaagagaggtttctgcagagattagattttaccaac	718
Db 76308	TCACAGACGGCATGGATACTGACAGGTCGTGAGTTTGTGGACAACGTTGTGGAGGTATCAT	76367
QY 719	acagctggctgcagaccccgcccttggtagaagagcccttgcacgagattccaggttg	778
Db 76368	GTTCCTTGGATTCCAGCCCGGATACAGTACGAGAGGCTCTTAAAAATGCCAAGACGTCC	76427
QY 779	accacagtggagattgtggaactggcgaaggtgcattctcctggaagagcatctct	838
Db 76428	ATCCACCGCGGAGATCTTG---TTTGAAGAATTTTCGCCCTGGAAGTCTCATCTGT	76484
QY 839	accacctggaaatctgggctgtccctccagtgcccatcttcttggttatctacactgacc	898
Db 76485	TCGATCTTGAA---AAGGACTACAAGGTGGAGGGTGTGCCAAGCTGGTGGTCTTCAACA	76541
QY 899	aggctgacagtgccgaataacagtggtgtgcccaagagagcccaactcattccaaagcggc	958
Db 76542	CGGCAATAGTTGGCGGAGTGGCGGAGTTCAGTGACCTGGCAGCTATCTGTGGCGCGCA	76601
QY 959	tgccctgcagagagccatggcggtgttcgggacagagccctggaccaggtcagtgga	1018
Db 76602	AATTTCTGCCCACTCCTTGGCGGAGTGTATGACGATGAACATATGCGATAAGCGAGCA	76661
QY 1019	tcctggctgcatcttcgtccatgcgaagcggttcattggcggctcaccgcacccagaggg	1078
Db 76662	TCAAAGATCTCGAAATTCATCCATCACACAAGATTATTGGCGGGGCAAAACCGAAGG	76721
QY 1079	gtgccttgagcagtgcc 1095	
Db 76722	CTGCCATGTGTGGCC 76738	
RESULT 3		
AC008254		
LOCUS	AC008254 95628 bp DNA HTG 17-FEB-2000	
DEFINITION	Drosophila melanogaster chromosome 3 clone BACR02G22 (D833) RPCI-98	
	02.G.22 map 85C-85C strain Y; cn bw sp, *** SEQUENCING IN PROGRESS	
	***, 71 unordered pieces.	
ACCESSION	AC008254	
VERSION	AC008254.5 GI:6984356	
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
	1 (bases 1 to 95628)	
	Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,	
	Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,	
	Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,	
	Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,	
	Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,	
	Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,	
	Richards, S., Sethi, H., Swirskas, R.R., Wan, K.H., Webster, D.,	
	Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.	
	Sequencing of Drosophila melanogaster	
TITLE	Unpublished	
JOURNAL	2 (bases 1 to 95628)	
REFERENCE	Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,	
AUTHORS	Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,	
	Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,	
	Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,	
	Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,	
	Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,	
	Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,	
	Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and	
	Rubin, G.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (30-JUL-1999) Drosophila Genome Center, Lawrence Berkeley	
	Laboratory, MS 64-121, Berkeley, CA 94720, USA	
COMMENT	On Feb 17, 2000 this sequence version replaced gi:6957906.	
	For further information about this sequence, including its location	
	and relationship to other sequences, please visit our sequence	
	archive Web site (http://www.fruitfly.org/sequence/) or send email	
	to bdgs@fruitfly.berkeley.edu . All contigs in this submission meet	
	the following cutoffs: length >= 200 bases.	
	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 71 contigs. The true order of the pieces	
	* is not known and their order in this sequence record is	
	* arbitrary. Gaps between the contigs are represented as	
	* runs of N, but the exact sizes of the gaps are unknown.	
	* This record will be updated with the finished sequence	
	* as soon as it is available and the accession number will	
	* be preserved.	
	* 1 1158: contig of 1158 bp in length	
	* 1159 1238: gap of unknown length	
	* 1239 2027: contig of 789 bp in length	

* 2028	2107: gap of unknown length	72102	
* 2108	2919: contig of 812 bp in length	72688	72687: contig of 586 bp in length
* 2920	2999: gap of unknown length	72768	72767: gap of unknown length
* 3000	3593: contig of 594 bp in length	73404	73403: contig of 636 bp in length
* 3594	3673: gap of unknown length	73484	73483: gap of unknown length
* 4216	4215: contig of 542 bp in length	74120	74119: contig of 636 bp in length
* 4296	4295: gap of unknown length	74200	74199: gap of unknown length
* 4939	4938: contig of 643 bp in length	74837	74836: contig of 637 bp in length
* 5019	5018: gap of unknown length	74917	74916: gap of unknown length
* 5686	5685: contig of 667 bp in length	75532	75531: contig of 615 bp in length
* 5766	5765: gap of unknown length	75612	75611: gap of unknown length
* 6418	6417: contig of 652 bp in length	76258	76257: contig of 646 bp in length
* 6498	6497: gap of unknown length	76338	76337: gap of unknown length
* 6988	6987: contig of 490 bp in length	76973	76972: contig of 635 bp in length
* 7068	7067: gap of unknown length	77052	77051: contig of 622 bp in length
* 7958	7957: contig of 890 bp in length	77532	77531: contig of 669 bp in length
* 8038	8037: gap of unknown length	77755	77754: gap of unknown length
* 8794	8793: contig of 756 bp in length	78424	78423: gap of unknown length
* 8874	8873: gap of unknown length	78504	78503: contig of 597 bp in length
* 10295	10294: contig of 1421 bp in length	79101	79100: gap of unknown length
* 10375	10374: gap of unknown length	79181	79180: contig of 229 bp in length
* 11600	11599: contig of 1225 bp in length	79410	79409: gap of unknown length
* 11680	11679: gap of unknown length	79490	79489: contig of 611 bp in length
* 13067	13066: contig of 1387 bp in length	80101	80100: gap of unknown length
* 13146	13145: gap of unknown length	80181	80180: contig of 615 bp in length
* 13147	13146: contig of 1208 bp in length	80796	80795: gap of unknown length
* 13355	13354: gap of unknown length	80876	80875: contig of 676 bp in length
* 14335	14334: contig of 1511 bp in length	81552	81551: gap of unknown length
* 15946	15945: gap of unknown length	81632	81631: contig of 791 bp in length
* 16026	16025: contig of 774 bp in length	82423	82422: gap of unknown length
* 16799	16798: gap of unknown length	82503	82502: contig of 596 bp in length
* 16800	16800: gap of unknown length	83199	83198: gap of unknown length
* 16880	16879: gap of unknown length	83279	83278: contig of 611 bp in length
* 18484	18483: contig of 1604 bp in length	83890	83889: gap of unknown length
* 18563	18562: gap of unknown length	83970	83969: contig of 591 bp in length
* 18564	18563: gap of unknown length	84561	84560: gap of unknown length
* 19520	19519: contig of 956 bp in length	84641	84640: contig of 623 bp in length
* 19599	19598: gap of unknown length	85264	85263: contig of 623 bp in length
* 19600	19599: gap of unknown length	85344	85343: gap of unknown length
* 20788	20787: contig of 1188 bp in length	86098	86097: contig of 754 bp in length
* 20867	20866: gap of unknown length	86178	86177: gap of unknown length
* 22258	22257: contig of 1391 bp in length	86721	86720: contig of 543 bp in length
* 22338	22337: gap of unknown length	86801	86800: gap of unknown length
* 24017	24016: contig of 1679 bp in length	87421	87420: contig of 620 bp in length
* 24097	24096: gap of unknown length	87501	87500: gap of unknown length
* 25669	25668: contig of 1572 bp in length	88123	88122: contig of 622 bp in length
* 25749	25748: gap of unknown length	88203	88202: gap of unknown length
* 25750	25750: contig of 1764 bp in length	88557	88556: contig of 654 bp in length
* 27514	27513: contig of 1764 bp in length	88936	88935: gap of unknown length
* 27594	27593: gap of unknown length	89372	89371: contig of 791 bp in length
* 28776	28775: contig of 1182 bp in length	89728	89727: gap of unknown length
* 28855	28854: gap of unknown length	89808	89807: gap of unknown length
* 30563	30562: contig of 1708 bp in length	90379	90378: contig of 571 bp in length
* 30643	30642: gap of unknown length	90459	90458: gap of unknown length
* 34088	34087: contig of 3445 bp in length	90991	90990: contig of 533 bp in length
* 34168	34167: gap of unknown length	90992	90991: gap of unknown length
* 35556	35555: contig of 1388 bp in length	91072	91071: contig of 659 bp in length
* 35636	35635: gap of unknown length	91731	91730: contig of 659 bp in length
* 37836	37835: contig of 2200 bp in length	91811	91810: gap of unknown length
* 37916	37915: gap of unknown length	92371	92370: contig of 560 bp in length
* 39435	39434: contig of 1519 bp in length	92451	92450: gap of unknown length
* 39515	39514: gap of unknown length	93164	93163: contig of 713 bp in length
* 43163	43162: contig of 3648 bp in length	93244	93243: gap of unknown length
* 43243	43242: gap of unknown length	94030	94029: contig of 786 bp in length
* 47050	47049: contig of 3807 bp in length	94110	94109: gap of unknown length
* 47130	47129: gap of unknown length	94897	94896: contig of 788 bp in length
* 51266	51265: contig of 4136 bp in length	94978	94977: gap of unknown length
* 51346	51345: gap of unknown length	95628	95627: contig of 651 bp in length.
* 51347	51346: contig of 7041 bp in length		
* 58388	58387: gap of unknown length		
* 58468	58467: gap of unknown length		
* 69717	69716: contig of 11249 bp in length		
* 70509	70508: gap of unknown length		
* 70510	70509: contig of 713 bp in length		
* 70590	70589: gap of unknown length		
* 71257	71256: contig of 667 bp in length		
* 71337	71336: gap of unknown length		
* 72022	72021: contig of 685 bp in length		
	72101: gap of unknown length		

FEATURES
source

Location/Qualifiers
1. 95628
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"
/map="85C-85C"
/clone="BACR02G22 (D833) RPCI-98 02.G.22"

```
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Query Match      8.3%; Score 97.8; DB 55; Length 95628;
Best Local Similarity 49.6%; Pred. No. 5.8e-14;
Matches 306; Conservative 0; Mismatches 307; Indels 4; Gaps 2;

QY 479 tgggaccctctatgacaagatgatgagaaactttgtggaggaggtggatgctgtggaca 538
Db 58687 TGCAACGGCCTTTATTTCAGATCTATAGGAATTTTATCAACGAACCTGGACGCAATTGATA. 58746

QY 539 atgggatctccagtgaggcagagggagcctcgatgacactgacacaccctcagtg 598
Db 58747 ATGGAGTGCCCATGTTTCGAGGGCGTTGAGCCTATATACAAAAATCTCCATCTATCT 58806

QY 599 cagagtgctgcactaaactcctacctgaaacaccccgacacacactgagcaggt 658
Db 58807 CCGCATTCGAACATGATCCGCTCTGGCAGGAACCTGGCGTAGACATCGAAGACCGAT 58866

QY 659 tcaagcgtgcaatggatctgttcaagagaggtttcttgcagagattagattttacacaac 718
Db 58867 TCAGACAGGCGATGATGACAGGTGCTGAGTTTGTGGACAACGTTGTGGAGGTATCAT 58926

QY 719 acagctggtgcagcccgccgcttggtagagagcccttgcacagcattccagtg 778
Db 58927 GTTCCTGGATTCCAGCCCGGATCACGTACGAGAGGCTCTTAAAAATGGCAAGACGTCC 58986

QY 779 accaagtggagattgtgaaactgcaagagtgcatctccctgaaagagcatctct 838
Db 58987 ATCCCAACGGGAAGATTCTT-GTGTGAAGAAATTTCTGCCCTGGAAGTCTCATCTGT 59045

QY 839 accacctggaatctggcgtgctccctccagtgccatcttcttggatctacactgacc 898
Db 59046 TCGATCTCTGAAGAGGATACAAAGTGAGGAGGTGTGCCAAGCTGGTCTTCAACACGG 59105

QY 899 agcctgacagtgccgaatacagtgtgtgcccagagagcccccactcattccaaagccgc 958
Db 59106 GCAATAG--TTGGCGAGTGGCGGAGTTCCAGTACGCTTGGCAGTATCTGGCGCGCA 59162

QY 959 tgccttcgcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1018
Db 59163 AATTCTTGCCCACTCTCTGGCGGAGTGTGAGCAGTAACTATGCGATGAAAGCGGAGCA 59222

QY 1019 tccctggtgcattcttgcctgcaagcagcagcagcagcagcagcagcagcagcagcagc 1078
Db 59223 TCAAGATCTCGAATTCATCCATCACACAGATTATTGGCGGGCAAAACCGAGGAGG 59282

QY 1079 gtgcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1095
Db 59283 CTGCCATGTTGTGGCC 59299

RESULT 4
CNS01G20/c 922 bp DNA STS 17-FEB-2000
LOCUS Anopheles gambiae STS SP6 end of clone 05N18 of NotreDamel library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
sequence tagged site.
ALI42585
ALI42585.1 GI:7000703
STS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Culicidae; Anopheles.
REFERENCE 1 (bases 1 to 922)
Genoscope.
DIRECT SUBMISSION
AUTHORS Direct Submission
TITLE Anopheles gambiae STS SP6 end of clone 05N18 of NotreDamel library
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 922)
REFERENCE 2 (bases 1 to 922)
Genoscope.
AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
```

```
Direct Submission
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
Location/Qualifiers
Source 1..922
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_lib="05N18"
/clone="NotreDamel"
/note="end : SP6"
BASE COUNT 219 a 271 c 227 g 203 t 2 others
ORIGIN

Query Match      8.3%; Score 97.6; DB 13; Length 922;
Best Local Similarity 53.5%; Pred. No. 7.6e-14;
Matches 246; Conservative 1; Mismatches 210; Indels 3; Gaps 2;

QY 651 ggcagggttcaagcgtgcaatggatctgttcaagagaggtttcttgcagagattagattt 710
Db 907 GGAGCGTTTCGAGAAGGCCAAAGCGTACGKGGACTCGAGTTTGTGCACAAGGTGACGTA 848

QY 711 ctaccaacacagctggtgcagcccgccgcttggtagaagagcccttgcacagcatt 770
Db 847 CTACGCCAGCTGTGTGGTGGCGGCACGCGATATTGTCTCAAGGGGTGGCAACCGGAT 788

QY 771 ccaggttgacccaagtgagagattgtgaaactggcgaaggtgcattccctggaagga 830
Db 787 GGCTTTCACAGTCGCGGGAGATTATTAGAGCT--CGAACCAACCGTGCCTGGTGAAGGA 730

QY 831 gcatctctacacactggaatctggctgtccctccagtgccatcttcttggatctcta 890
Db 729 GCATCTCTATCAGCTGCGAGCAGGAGCAGAACTCTGTCTGGCGCAGGCCAAGTACGTGATCTA 670

QY 891 cactgacacagctggaacagtggtggaataacagtggtgtgtcccaagagagcccaactatcca 950
Db 669 CTGCAATTAAGAAAACGATTGGCGCGTCATATGCGCGCTTGCAGCGCGCCAGCTTTGT 610

QY 951 aagcggctgcccctgcagagccatgcccgggtcttcgagacagccctgagcaggt 1010
Db 609 GTGTGCAAGTTCTTGCCCAATCTGTGGCGGAGTAGCGGACGAGGAGCTAGAGAAGGT 550

QY 1011 cagtggtgacccctggtgcatcttctgctccatgcaagcggtcttcattggcggtaccgcac 1070
Db 549 GTCGGTATTGGCGGAT-CAACTTTTGGCCACCAGACCGGCTTCATTGGGGTAACAAAC 491

QY 1071 ccgagaggggtgcttgagcagtgccccgtgcaaccttggcc 1110
Db 490 GCGGAAGGTGCGCTCAAGATGGCAATTAGCTAGTTTGGC 451

RESULT 5
AC017792
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC017792
VERSION AC017792.1 GI:6553398
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 123579)
Adams,M. and Venter,J.C.
AUTHORS Direct Submission
TITLE Drosophila melanogaster HTG
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
```

Rockville, MD, USA
This sequence was identified as CDM:10212105 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

ORGANISM Botryotinia fuckeliana
Eukaryota; Fungi; Ascomycota; Euascomycetes; Discomycetes;
Leotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 720)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
78026 Versailles, France
2 (bases 1 to 720)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage ;
CP 5706 91057 EVRY cedex - FRANCE (E-mail : segrefgenoscope.cns.fr
Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this
project was created using a Botrytis cinerea strain which was grown
under conditions of nitrogen deprivation, which is the normal
situation for B. cinerea during its development on its host plant.
The library was produced in an oriented direction, in the pBSII
vector.

FEATURES Location/Qualifiers
source 1..720
/organism="Botryotinia fuckeliana"
/strain="T4"
/db_xref="taxon:40559"
/note="Genoscope sequence ID : W67F011"
BASE COUNT 214 a 166 c 154 g 186 t
ORIGIN

Query Match 7.3%; Score 86.2; DB 8; Length 720;
Best Local Similarity 53.6%; Pred. No. 4.8e-11;
Matches 215; Conservative 0; Mismatches 168; Indels 18; Gaps 1;

QY 148 atgggagcacaatggcaaccttcacatgcacagcagcactggcgcactgcttcgc 207
DB 107 ATGGGTACTCATATGGCCCAATTTTCCACGAGATGAAGCTTGGCTGTTTACNCGTTCGC 166
QY 208 ctctccggaggtaccggatgcagagattgtgcggaccctggatcccgaaaaactgcct 267
DB 167 CTCTCTCTTACTTATCAATCTTCAGAGCTCATTCGAACTCGGATGCCAAACTTCTAGAG 226
QY 268 tctctgacatcgtggtgagcgtggggggcagtagcacacctcggagacacgcatgac 327
DB 227 ACTTGGCATACCGCTGGTGTGATGTGGGAGGTGAATACACGACGAACATPACAGATATGAT 286
QY 328 catcacagaggtcttccacagagaccatgagctccctgtccctgggagcctggcag 387
DB 287 CACCATCAAGTACTTTCGATACCCACATTTCCCAATC-----GTCT 328
QY 388 accaagctgagcagtgctggggaactcatctctgacactcgggacacaaagtctgcccag 447
DB 329 ACCAAGCTCTCTTCTCGGGGTTAGTGTATATGCATACGCGCAAGCGCATATCGCACA 388
QY 448 ttgctggcactagtcaagagacagcagcattggtggccacctctatgacaagtatgatgag 507
DB 389 CATCTAGTGTGCCGAAGATCGGGAAGAGTTCGCTTATCTGGAGAAGATTTACGAA 448
QY 508 aactttgtgagaggtgagctgctggacaatgggatctc 548
DB 449 AGCTTTATTGAAGCACTTGATGCTCAGGATAACGGTATTTC 489

RESULT 7
SPAC694/c 18700 bp DNA PLN 02-FEB-2000
LOCUS S.pombe chromosome I cosmid c694.
DEFINITION AL138666
ACCESSION AL138666.1 GI:6901195
VERSION 408 ribosomal protein s25 (s31); conserved hypothetical; DEAD/DEAH
KEYWORDS box helicase; helicase; rps25; S1 RNA binding domain; transcription
initiation protein; zinc-finger.
SOURCE fission yeast.


```
CDS
/ gene="SPAC694.04c"
complement(join(8981..9900,10244..10298))
/ gene="SPAC694.04c"
/ note="SPAC694.04c, len:324, SIMILARITY:Saccharomyces
cerevisiae, YEF6_YEAST, hypothetical 38.2 kd protein in
bem2-ncbl intergenic region., (338 aa), fasta scores: opt:
1016, E():0, (49.3% identity in 306 aa)"
/codon_start=1
/label="SPAC694.04c"
/product="conserved hypothetical protein"
/protein_id="CAB71842.1"
/db_xref="GI:6901199"
/translation="MNNLVKTIATSGTHFADELAVYMLRLDRFSGAQIVRSRDPQV
LSDCIIVDVGKYGDIKFYDHHQREFNDTSPKYSRLSSAGLIYKHFGEVYHAVL
POLKINEODLETLYEKVQSFVEGLDNDNGISAYPAGLKPSFKAAASLPEMVSFLP
ANNSKODDOTLECFQKASDLGMFWFVSVALSWLPARTLAREAILKAKDSPIL
IVDQPPMKGHFLDEKEGLIQNKYAIYSDGKAWRQVAVSIDTSETCRLPLPEW
RGIRDEKLSELGIPCCIFVHASGFIGGNQTFEGALEMARKALDPQN"
complement(9901..9920)
/misc_feature
/ gene="SPAC694.04c"
/ note="ctaattataaatatttag, splice branch and acceptor"
complement(10238..10243)
/misc_feature
/ gene="SPAC694.04c"
/ note="gtgaagt, splice donor sequence"
complement(10661..11033)
/gene="rps25"
complement(join(10661..10686,10740..10980,11031..11033))
CDS
/ gene="rps25"
complement(join(10661..10686,10740..10980,11031..11033))
/ gene="rps25"
/ note="SPAC694.05c, len:89"
/codon_start=1
/label="rps25"
/product="40s ribosomal protein s25 (s31)"
/protein_id="CAB71843.1"
/db_xref="GI:6901200"
/translation="MAPKKKSKGKVKDKAQHATVFDKSIIDRINKVPAFKFISVSV
LVDRMKINGSLARIAIRDLAERGVIOKVDQSKQAIYTRVPAATA"
complement(10687..10739)
/intron
/ gene="rps25"
/ note="confirmed intron"
complement(10687..10702)
/misc_feature
/ gene="rps25"
/ note="ctaacaatttttag, splice branch and acceptor"
complement(10734..10739)
/misc_feature
/ gene="rps25"
/ note="gtgaagt, splice donor sequence"
complement(10981..10995)
Query Match 6.5%; Score 76.4; DB 8; Length 18700;
Best Local Similarity 57.2%; Pred. No. 1.1e-08;
Matches 159; Conservative 0; Mismatches 116; Indels 3; Gaps 1;
Qy 817 tgtccctgaaggagcattctaccacctggaattctggctgtccctccagtgcccatc 876
Db 9286 TTTCCCTGGAAGGGTCATTATTGATATGAAAAGAACTTGGCATTGAGAACCAATTT 9227
Qy 877 tctttgttatctacacgacagctggacagtgagcgaatacagtggtgtgcccacagag 936
Db 9226 AAATACGCTATCTACTCTG--ATGGGAAGCTTGGAGTTCAGGCTGTTCTATTGAT 9170
Qy 937 cccactcattccaaagccgctgcctccctgcagagccatggcgggggtcctcgggagcag 996
Db 9169 CCCACTCTTTTACTTGTGCGTTACCACTACCAGAACCATGGAGGAATTCGTGATGA 9110
Qy 997 gccctggaccaggtcagtgaggatccctgctgctatcttcgtccctcgaacgagcgttcatt 1056
Db 9109 AAGCTTTCGAGTTAACTGGCAFTTCAGGTTGCATCTTTGTCATGCTCCGGTTTCATT 9050
Qy 1057 ggcgggtcaccgcaaccgagaggggtgccttgagcagtcgc 1094
Db 9049 GGTGGTAATCAACACTTTTCAGGGTCTCTGGAAATGGC 9012
```

```
CEC27H6/C
LOCUS
DEFINITION
Caenorhabditis elegans cosmid C27H6, complete sequence.
ACCESSION
281042
VERSION
281042.1
KEYWORDS
HTG; yeast hypothetical 50.5 KD protein like; Yeast hypothetical
protein YEF6 like.
SOURCE
Caenorhabditis elegans.
ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1. (bases 1 to 24775)
Genome sequencing of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
The C.elegans Sequencing Consortium.
Erratum: [[published errata appear in Science 1999 Jan
1:283(5398):35 and 1999 Mar 26;283(5410):2103]]
2. (bases 1 to 24775)
Gardner, A.E.
Direct Submission
Submitted (21-OCT-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 6310, USA. E-mail:
jes@sanger.ac.uk or rw@nematoe.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormacsc&class=Sequence&object=C27H6
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone C27H6.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone C27H6 is at 33379 in
sequence 873976.
The true right end of clone C27H6 is at 10790 in
sequence 88113.
The true left end of clone K08H10 is at 24675 in this sequence. The
true right end of clone T07C12 is at 104 in this sequence. The
start of this sequence (1..104) overlaps with the end of sequence
273976.
The end of this sequence (24672..24775) overlaps with the start of
sequence 283113.
FEATURES
Location/Qualifiers
1..24775
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="C27H6"
complement(4344..13781)
/ gene="C27H6.1"
complement(join(4344..4473,4520..4665,4959..5591,
5849..6913,7085..7834,8078..8205,8256..9164,9337..9619,
9737..9781,10254..10405,12673..12739,13116..13781))
/ note="Weak similarity to C elegans clathrin coat assembly
protein AP47 (SW:AP47_CAPEL); cDNA EST EMBL:Z14859 comes
from this gene; cDNA EST EMBL:D64772 comes from this gene;
cDNA EST EMBL:D34082 comes from this gene; cDNA EST
```

```
CEC27H6
24775 bp
DNA
INV
14-DEC-1999
Caenorhabditis elegans cosmid C27H6, complete sequence.
281042
GI:1627626
HTG; yeast hypothetical 50.5 KD protein like; Yeast hypothetical
protein YEF6 like.
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1. (bases 1 to 24775)
Genome sequencing of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
The C.elegans Sequencing Consortium.
Erratum: [[published errata appear in Science 1999 Jan
1:283(5398):35 and 1999 Mar 26;283(5410):2103]]
2. (bases 1 to 24775)
Gardner, A.E.
Direct Submission
Submitted (21-OCT-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 6310, USA. E-mail:
jes@sanger.ac.uk or rw@nematoe.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormacsc&class=Sequence&object=C27H6
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone C27H6.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone C27H6 is at 33379 in
sequence 873976.
The true right end of clone C27H6 is at 10790 in
sequence 88113.
The true left end of clone K08H10 is at 24675 in this sequence. The
true right end of clone T07C12 is at 104 in this sequence. The
start of this sequence (1..104) overlaps with the end of sequence
273976.
The end of this sequence (24672..24775) overlaps with the start of
sequence 283113.
Location/Qualifiers
1..24775
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="C27H6"
complement(4344..13781)
/ gene="C27H6.1"
complement(join(4344..4473,4520..4665,4959..5591,
5849..6913,7085..7834,8078..8205,8256..9164,9337..9619,
9737..9781,10254..10405,12673..12739,13116..13781))
/ note="Weak similarity to C elegans clathrin coat assembly
protein AP47 (SW:AP47_CAPEL); cDNA EST EMBL:Z14859 comes
from this gene; cDNA EST EMBL:D64772 comes from this gene;
cDNA EST EMBL:D34082 comes from this gene; cDNA EST
```



```
/protein_id="AAF16577.1"
/db_xref="GI:6553912"
/translation="MEVEYLVGVMAAVALSLKTKNDRINTSSSFNSKNVLL
VFSIMAGWLOGPVYIYLYSYGFGKGIDGQIFAGFGSSMLFGTIYVSLADKQGRK
RACVTICVILSKSPQKVLKVGRLGDIATSLFSAFESLAEHKNRNEPQ
QWLSSTFSKVLGNGLVAILSGLFGNLVDTSFGPVAFDAACFLAIGMAILGT
WSENGFDSDSLDTQFVAAIAIASDEKIALGAIQSLFEASMTFFVFLWTPALSP
NDEEIPGHVFATFMLASGLSSLAARLSRSLRVNMQIVFLVSAASLLLPITS
VLVTPSKVDEGLSTSSITQLLGFCVFCFSCVGIFWPISIMKMSQYIPEEARSTINNF
RVPLNIFVICVLVNDAPFITMFGMCSIFLVSILQRLRLMVLVISEKPKAEDWSPMKE
RNSVDPLPL"
complement(join(<9170, .9394, 9488, .9559, 9692, .9811,
10121, .10312, 10653, .10751, 10822, .10968, 11209, .11431))
/genes="TIG12.2"
complement(<9170, .11431)
/genes="TIG12.2"
/notes="similar to hypothetical protein (Yer156cp)
GB:6321004 [Saccharomyces cerevisiae]"
complement(join(9170, .9394, 9488, .9559, 9692, .9811,
10121, .10312, 10653, .10751, 10822, .10968, 11209, .11418))
/genes="TIG12.2"
/notes="unknown protein"
/codon_start=1
/protein_id="AAF16578.1"
/db_xref="GI:6553913"
/translation="MFGSRLGCCSRIRNRSFLKRSNFRASFSTRKVTGHTGPHCD
EALACILRRNRFSDAOIVRTDHOVLKLDLDAALDVGVDYDQSPERYDHHOKGSEV
FGLGNTKLSAGLYKHKHGLIISKELOEQRHPDVFRLFLAYKNFIEAVDALDNG
IHQYDQPPRVNNTSLGRIGRLMLDIEPDQSAKEDAFHRAELAGSEFLFCV
HFHAKSLPVIEMCLAKRYDIDSSGEIMKLSKQCPWKLHIFELEEMKIDPPIKY
VLYQDRSNRIRIQAVSVSPERFESRKALPLAWRGLKEKLSSESIIPRCVVFHMSGF
IGANOTYEGALAMARASLMA"
<12217, >12837
/genes="TIG12.3"
<12217, >12837
/genes="TIG12.3"
/notes="predicted by genscan"
12217, .12837
/genes="TIG12.3"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF16579.1"
/db_xref="GI:6553914"
/translation="MILPKLSLEHIKTPPEOMAYVHLLMSLLIFITFKKSISSVVP
DPPTQNLIRICKQIDFECQQTITSQIVRQTSIKTKAKTLTAARAWINATQILDNI
EDSLPKADKRDQACRAKAYLVDALHONALAYLDRFRMRQALALWNIS
MCRSTSPFHTPMYANWNNKMLTDIAIYAGKILSPPLPTTKSKSP"
complement(join(<14136, .14716, 14922, >15366))
/genes="TIG12.4"
complement(<14136, >15366)
/genes="TIG12.4"
/notes="similar to cysteine proteinase GB:AAA50755 [Alnus
glutinosal; contains pfam profile: PF00112 papain family
cysteine protease]"
complement(join(14136, .14716, 14922, .15366))
/genes="TIG12.4"
/codon_start=1
/product="putative cysteine proteinase"
/protein_id="AAF16580.1"
/db_xref="GI:6553915"
/translation="MTSIVFFLLAILLSRTSGVTSRGGIIFEASAVEKHEQWMSRNR
VYSDSEKTSRFEIFNNLKVESINNTNKTVLDVNEFSDLTDEFKARYGLVVP
EGMTRISTDHSHTSVFRVNGETGESMDIOEGAVTSVKHQOQCCGCAWFAVAAY
EGMTKIANGLSESQQLDSCSTENNGCGGIMWKAIFYIKENQGITTEDNYPQGA
QQTCESNHIAAATISGYETVPQNDKALKAVSQQVSAIEGSGYEFTHYSGGIFNG
ECGTQLTHAVTIYGVGSEBEGIKYLLKNWSGESGNGYMRIMRDVDSPOQMGCLAS
LAYPVA"
complement(join(<18455, .19721, 19777, .19830, 19970, .20019,
20116, .20251, 20481, >20638))
/genes="TIG12.5"
complement(<18455, >20638)
/genes="TIG12.5"
/notes="contains Pfam profile: PF00566 TBC domain (implies
```

```
that these domains are GTPase activator proteins of
Rab-like small GTPases)"
complement(join(18455, .19721, 19777, .19830, 19970, .20019,
20116, .20251, 20481, .20638))
/genes="TIG12.5"
/codon_start=1
/product="putative GTPase activator protein of Rab-like
small GTPases"
/protein_id="AAF16581.1"
/db_xref="GI:6553916"
/translation="MKALRRIQTTSSNPPSLSPSSSSSSSSSIQIRSAFLVVAS
SPASCDRPLKSPWRSRKRKRLRPQMKRFFPDGRLRNGCVLLKLRVGRGIEP
STRLEYWPLLLGLYFNSKEERTVIRNRSSFFDSIAHRFCYKCRKRELRRQCKR
LOKHNGTRKLYNGSETIOEDYWPVQPTDSSCSDSEVVSARESLSDDEDITDICYM
SEWCVRERDGCSSRRITNATISTLNSSESDSDSEFVQVFOSSATPDVNSYTP
ASSIPRTEDEFVWORIRLDRADRADEWTSYSPQAVISEVQACRAAEVGLKDIY
HLEPYIFQAARLAVLAEYALYDPDGYCQGMDDLSPILSVIPDHEVFWCFVGFPM
KARHNRIDVEGIRRLQINIVSKIISKDQLYHLEKQAEFCFFYRMVYVWFRE
LTIDQTLCLWEVNWADQAAIRAGMGKSAMSRIRQRAPPTDLDLYAIAASVLQRRKRI
IERYSMDLETRECQSMAGLDVWKLDDAHLVLTHTKIEHSFS"
20486, .20608
/repeat_region
/rpt_family="(GAA)n"
complement(20824, .20921)
/rpt_family="(GAA)n"
complement(join(23851, .24472, 24683, .24818, 24924, >25075))
/genes="TIG12.6"
complement(23851, >25075)
/genes="TIG12.6"
/notes="similar to 6-phosphogluconolactonase GB:CAB57866
[ Homo sapiens]"
complement(join(23981, .24472, 24683, .24818, 24924, .25075))
/genes="TIG12.6"
/codon_start=1
/product="putative 6-phosphogluconolactonase"
/protein_id="AAF16582.1"
/db_xref="GI:6553917"
/translation="MAPVKRRVKTNNMAVELAKYTADELSSKFKERGVYTVVLSGG
DLIALWKLLEAPYIDSIEMSKWHIFWDERVCAWDHADSNYKLAYDGLSKVPVPAE
NIYALDNGLAGENAEALAEERYECLKQKVNQNIIRYKSSGPPQDQLLLGMPDGH
MASLFPQHAQINEKVRWVTSITSDSPPPSKRITLTLPVINCASINWAVCDKQKQADV
AAALNHTKDLPAGRLTADVWVWFLDQAAASKLPHGWCSTL"
complement(join(<25746, .25886, 25974, .26141, 26245, .26307,
26411, .26515, 26708, .26817, 26899, .27195, 27302, .27380,
Query Match 4.68; Score 53.6; DB 8; Length 80367;
Best Local Similarity 53.94; Pred. No. 0.0041;
Matches 110; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 904 ggacagtggcgaatacacagtgtgtgcccaggagcgcctcattcccaagccggtgccc 963
Db 9382 GAAACATGGAGAAATACAGCAGTTTCGTTTACACAGAGAGGTTTGAGAGCCGTAAAGCT 9323
QY 964 ctgccagagccatggcgggggtcttcgggacgagcgccctggaccaggtcagtgggatccct 1023
Db 9322 TTACCATTAGCATGAGAGGTTTAGAAAAAGGAGAGCTCTCAGAGAGAAAGCTCAATTCCA 9263
QY 1024 ggtcgatcttcgtccatgcgaagcggtcttcattggcggtaccaccgagaggggtgcc 1083
Db 9262 CGCTGTGTTTTTGTGCATATGATGAGTGGTTTCATTGTTGTCGTAACCAACCTATCAAGGTGCC 9203
QY 1084 ttgagcatggccctgcacattg 1107
Db 9202 TTACCAATGGCAAGAGCTTCCTTG 9179
RESULT 11
ATF7K15/c
LOCUS ATF7K15 129757 bp DNA PLN 19-JAN-2000
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone F2K15.
ACCESSION AL132956
VERSION AL132956.2 GI:6723384
KEYWORDS thale cress.
```


* 9632	9711: gap of unknown length	* 49831	51178: contig of 1348 bp in length
* 9712	10555: contig of 844 bp in length	* 51179	51258: gap of unknown length
* 10556	10635: gap of unknown length	* 51259	52725: contig of 1467 bp in length
* 10636	11493: contig of 858 bp in length	* 52726	52805: gap of unknown length
* 11494	11573: gap of unknown length	* 52806	53933: contig of 1148 bp in length
* 11574	12500: contig of 927 bp in length	* 53954	54033: gap of unknown length
* 12501	12580: gap of unknown length	* 54034	54756: contig of 723 bp in length
* 12581	13656: contig of 1076 bp in length	* 54757	54836: gap of unknown length
* 13657	13736: gap of unknown length	* 54837	57336: contig of 2500 bp in length
* 14428	14428: contig of 692 bp in length	* 57337	57416: gap of unknown length
* 14508	14508: gap of unknown length	* 57417	59460: contig of 2044 bp in length
* 15666	15666: contig of 1158 bp in length	* 59461	59540: gap of unknown length
* 15747	15746: gap of unknown length	* 59541	61545: contig of 2005 bp in length
* 15971	15971: contig of 225 bp in length	* 61546	61625: gap of unknown length
* 16051	16051: gap of unknown length	* 61626	63058: contig of 1433 bp in length
* 16701	16701: contig of 650 bp in length	* 63059	63138: gap of unknown length
* 16781	16781: gap of unknown length	* 63139	64310: contig of 1172 bp in length
* 17592	17592: contig of 811 bp in length	* 64311	64390: gap of unknown length
* 17672	17672: gap of unknown length	* 64391	65860: contig of 1470 bp in length
* 17993	18300: contig of 628 bp in length	* 65861	65940: gap of unknown length
* 18301	18380: gap of unknown length	* 65941	67531: contig of 1591 bp in length
* 18381	19002: contig of 622 bp in length	* 67532	67611: gap of unknown length
* 19003	19082: gap of unknown length	* 67612	68756: contig of 1145 bp in length
* 19083	19709: contig of 627 bp in length	* 68757	68836: gap of unknown length
* 19710	19789: gap of unknown length	* 68837	70478: contig of 1642 bp in length
* 19790	20451: contig of 662 bp in length	* 70479	70558: gap of unknown length
* 20452	20531: gap of unknown length	* 70559	71866: contig of 1308 bp in length
* 20532	21255: contig of 724 bp in length	* 71867	71946: gap of unknown length
* 21256	21335: gap of unknown length	* 71947	74543: contig of 2597 bp in length
* 21336	21931: contig of 596 bp in length	* 74544	74623: gap of unknown length
* 21932	22011: gap of unknown length	* 74624	74664: contig of 1841 bp in length
* 22012	22750: contig of 739 bp in length	* 74665	76544: gap of unknown length
* 22751	22830: gap of unknown length	* 76545	78221: contig of 1677 bp in length
* 22831	23628: contig of 796 bp in length	* 78222	78301: gap of unknown length
* 23627	23706: gap of unknown length	* 78302	79825: contig of 1524 bp in length
* 23707	24763: contig of 1057 bp in length	* 79826	79905: gap of unknown length
* 24764	24843: gap of unknown length	* 79906	81889: contig of 1984 bp in length
* 24844	26231: contig of 1388 bp in length	* 81890	81969: gap of unknown length
* 26232	26311: gap of unknown length	* 81970	83864: contig of 1895 bp in length
* 26312	26933: contig of 622 bp in length	* 83865	83944: gap of unknown length
* 26934	27013: gap of unknown length	* 83945	86590: contig of 2646 bp in length
* 27014	28246: contig of 1233 bp in length	* 86591	86670: gap of unknown length
* 28247	28326: gap of unknown length	* 86671	88955: contig of 2285 bp in length
* 28327	29732: contig of 1406 bp in length	* 88956	89035: gap of unknown length
* 29733	29812: gap of unknown length	* 89036	92254: contig of 3219 bp in length
* 29813	30738: contig of 926 bp in length	* 92255	92334: gap of unknown length
* 30739	30818: gap of unknown length	* 92335	94954: contig of 2619 bp in length
* 30819	31683: contig of 865 bp in length	* 94954	95033: gap of unknown length
* 31684	31763: gap of unknown length	* 95034	97732: contig of 2699 bp in length
* 31764	33108: contig of 1345 bp in length	* 97733	97812: gap of unknown length
* 33109	33188: gap of unknown length	* 97813	101160: contig of 3348 bp in length
* 33189	35225: contig of 2037 bp in length	* 101161	101240: gap of unknown length
* 35226	35305: gap of unknown length	* 101241	103957: contig of 2717 bp in length
* 35306	36646: contig of 1341 bp in length		
* 36647	36726: gap of unknown length		
* 36727	37651: contig of 925 bp in length		
* 37652	37731: gap of unknown length		
* 37732	38821: contig of 1090 bp in length		
* 38822	38901: gap of unknown length		
* 38902	40239: contig of 1338 bp in length		
* 40240	40319: gap of unknown length		
* 40320	41306: contig of 987 bp in length		
* 41307	41386: gap of unknown length		
* 41387	42659: contig of 1273 bp in length		
* 42660	42739: gap of unknown length		
* 42740	44063: contig of 1324 bp in length		
* 44064	44143: gap of unknown length		
* 44144	45605: contig of 1462 bp in length		
* 45606	45685: gap of unknown length		
* 45686	47472: contig of 1787 bp in length		
* 47473	47552: gap of unknown length		
* 47553	49750: contig of 2198 bp in length		
* 49751	49830: gap of unknown length		
<hr/>			
Query Match 4.5%; Score 52.6; DB 41; Length 194613;			
Best Local Similarity 57.1%; Pred. No. 0.007;			
Matches 117; Conservative 0; Mismatches 84; Indels 4; Gaps 1;			
QY 137	caccgcccgcgaatcgggaagcacaatggcacccttcacatgcgacgagcgactggcatgcg 196		
DB 158659	CTCCCTTTGGATCGGCACACACAGCGGCACTTTTCTACGCGACGAGTGGTTCGCATGTT 158718		
QY 197	cactgttcctccctccgcggagatcaggatgtgcggaccgccggtatcccg 256		
DB 158719	TTATGCTCAAGCAATTTGGATGAGTACAAAACCGACAGATTTTCCGACGCGGCAAAAC 158778		
QY 257	aaaaactcgccttcc---tgtgacatcgtgtgtgagcgtggggggcgagtcagaccctcgg 312		
DB 158779	AAGCCTTTGAAGGAGAAATGCCAAATAATTTGCGACGTGGCGCGCTTTATGATCATGCC 158838		
QY 313	agacaccgatatgaccataccaga 337		
DB 158839	AAGAAATTTGATCAGATCACCACCAAA 158863		

Db	12146	AATGTGAGTACGCTCTGTGTAGTAGGAGCGCGTACTCTACAGAGCATAAAGCGTTTCGATC 12087	
Qy	329	atcacagaggtctttcacagagacccatgag 359	
Db	12086	ACCACCAAGTTCTTACACAGGATCTTGGAG 12056	
RESULT 14			
CEK08H10/c		DNA	INV 26-JAN-2000
LOCUS		Caenorhabditis elegans cosmid K08H10, complete sequence.	
DEFINITION		283113	
ACCESSION		283113	
VERSION		283113.1 GI:3217648	
KEYWORDS		HTG.	
SOURCE		Caenorhabditis elegans.	
ORGANISM		Caenorhabditis elegans	
REFERENCE		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.	
AUTHORS		1 (bases 1 to 39339)	
TITLE		none.	
JOURNAL		Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium	
MEDLINE		Science 282 (5396), 2012-2018 (1998)	
REMARK		99069613	
REFERENCE		Erratum: [[published errata appear in Science 1999 Jan 1; 283(5398):35 and 1999 Mar 26; 283(5410):2103 and 1999 Sep 3; 285(5433):1493]]	
AUTHORS		2 (bases 1 to 39339)	
TITLE		Gardner, A.E.	
JOURNAL		Direct Submission	
COMMENT		Submitted (27-NOV-1996) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rv@nematode.wustl.edu	
		On Jun 13, 1998 this sequence version replaced gi:1695070.	
		Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.	
		For a graphical representation of this sequence and its analysis see: "	
		http://webcace.sanger.ac.uk/cgi-bin/display?db=wormacesclass=Sequence object=K08H10	
		Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.	
		IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.	
		This sequence is the entire insert of clone K08H10. The true left end of clone T0688 is at 32990 in this sequence. The true right end of clone C27H6 is at 10790 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 281042.	
		The end of this sequence (39233..39339) overlaps with the start of sequence 273975.	
FEATURES		Location/Qualifiers	
source		1..39339	
		/organism="Caenorhabditis elegans"	
		/db_xref="taxon:6239"	
		/chromosome="V"	
		/clone="K08H10"	
		complement(1035..4606)	
gene		/gene="K08H10.7"	
		complement(join(1035..1256,1333..1815,1877..2485,2915..3296,3343..3559,3607..3785,3840..3895,3940..4252,4295..4494,4550..4606))	
CDS		/gene="K08H10.7"	
		/note="cDNA EST yk296b10.5 comes from this gene; cDNA EST yk296b10.3 comes from this gene; cDNA EST EMBL:D27223 comes from this gene; cDNA EST EMBL:D27222 comes from this	

```

gene; cDNA EST EMBL:D27221 comes from this gene; cDNA EST
EMBL:D27220 comes from this gene; cDNA EST EMBL:D32492
comes from this gene; cDNA EST EMBL:D3146 comes from this
gene; cDNA EST EMBL:D33093 comes from this gene; cDNA EST
EMBL:D35838 comes from this gene; cDNA EST yk2004.3 comes
from this gene; cDNA EST yk2004.5 comes from this gene;
cDNA EST yk214e12.3 comes from this gene; cDNA EST
yk214e12.5 comes from this gene; cDNA EST yk218f8.3 comes
from this gene; cDNA EST yk218f8.5 comes from this gene;
cDNA EST yk420e2.3 comes from this gene; cDNA EST
yk420e2.5 comes from this gene; cDNA EST EMBL:Z14900 comes
from this gene; cDNA EST EMBL:M69235 comes from this gene"
/codon_start=1
/protein_id="CAB05546.1"
/db_xref="GI:3924822"
/db_xref="SPTREMBL:Q9XU82"
/translation="MSSNPELEKFGYRHSLDPEMKWLARPCKDCDKYEYKVLILY
NMFKSKYIDREYIEYKMTKEVLRNRPAGFPFPAKTEIPIDRAKLWQHURHEKK
QTDILEDVFEKEDTVSVYRLNVTYKMLVSEKVKYKDKSEKDEKLEKKILYTM
LTYRKHFHFNRENPEKDEANRSYKFLNVTQKRYAPFYNEEKVQFAKNFYD
NSILRVPSFDPNRPQSLVAPRIEAWFGIYIGIKELFDEGPVLAIVDKLFYN
APKMSLLDYLIVDQSCNDVYKDLTKLMAGKMTIRQAARPRIRLLENKLKCA
EYWDNEMSLRTEHLFLDLCRENSLYVYKTKSDRGNAKYDTTLFKLIENKAFI
EPHPLPVKVSAGAYVPMHELVHEKPYQRKNRIDLVMQDFLKRATKRPDHYKE
NTLKMKELDLFSSEELNFFYKNCITLNTIGREFIAATEAKNMFRLPDEQKVLMP
ISGRQLNAYGVKHYCDHTYGVANQHTSETVTKALASRHEKSKRIFYQIALKINA
KLGGINOELDWSEIAEISPEEKERRKTMTPLTVGVGDVTHPTSYSGIDYSIAAVASI
NPGGTIVRNKIVTOECSRPGERAVANGRETDILEAKFKVLLREFAEANDNRPAHIV
VYRDGVSEMLRVSHDELRSLSKSEVKQFMSERGDPEPKYTFYIVQKRNRLLRR
MEKRDVANKDITPAETDVAVAQKQWEDKESKETGIVNPNSSFTYDVKLIVSKYR
DFFLASHGVGLTSPRGHTVMYDDKMSQDEYVMTYGLAFSLARCRKPSISLPVPH
YAHLSCEKAKELYRKYEHYIGDYAQPRTRHMEHEFLQTNVYKPGMSFA"
4899..5698
/genes="K08H10.9"
join(4899..4994,5051..5122,5167..5268,5318..5395,
5444..5584,5630..5698)
/genes="K08H10.9"
/notes="cDNA EST yk491a1.5 comes from this gene; cDNA EST
yk491a1.3 comes from this gene"
/codon_start=1
/protein_id="CAB05547.1"
/db_xref="GI:3924823"
/db_xref="SPTREMBL:Q9XU81"
/translation="MAFPDIDILHEIRLSTSEKDRCAKEKFGTIIDNLGPADIE
KYESVOKAPFLANAKLESIGFRVGRQLVKESPKLVEIVKFIKDFWSS
VTEGKQVDNRTAHQGVVVDQGRFTTLRSPFEGDQVKSSEYFLAPCGLLRGALGL
NRAIVPTIESVPSVKFHIQVQK"
complement(6724..8875)
/genes="K08H10.6"
complement(join(6724..6781,6827..6943,6995..7242,
7294..7435,7684..8083,8127..8425,8700..8875))
/genes="K08H10.6"
/notes="similar to Sugar (and other) transporters"
/codon_start=1
/protein_id="CAB05545.1"
/db_xref="GI:3924821"
/db_xref="SPTREMBL:Q9XU83"
/translation="MKVLPLEARNLETPPHLTTRLLATSVVAVFAGGFHGYLLISAVNP
LSLILQOQVNDLQNRYSIELSSQLSLWSSLAGCLFICAMIGATISIFLQSIQPR
NSLLAAIQLFSAPVFGYLAISRLCELLPISRLISGLIAFAGISAQGFLEISPAK
FRGLTSLSLGACFLAALATGTGVLTGTVMQRYIFFETIPCVLHILVNITTFH
DSPYLLSLGKISGESAIKVVYGETCHKIKIMEDLRIIDNGVKESLRQILKDKGA
QALSIAVAFNSVAFSIVAIISPEFTFLQITGFSPESAVANSLSASFASIVSALLAA
IANDKIGRRPLTISLLVIALINIFMMSLVFLYDSKDSFLAWPFLALVLFVTFVESI
GIGPAAVFICAEALAPPTGISKMSQSYSTVSQFAGSFCTPIIYIQLVSEIGFAPMLFIV
PLTITAAVYFMYLPETKGPSVEIYGLLQ"
17960..20251
/genes="K08H10.4"
join(17960..18067,18287..18427,18475..18790,19256..19589,
19640..19976,20048..20251)
/genes="K08H10.4"
/notes="similar to GDAL/CD39 (nucleoside phosphatase)
family; cDNA EST yk400d11.5 comes from this gene; cDNA EST

```

```

yk242e9.3 comes from this gene; cDNA EST yk262c6.3 comes
from this gene; cDNA EST yk262c6.5 comes from this gene;
cDNA EST yk400d11.3 comes from this gene; cDNA EST
EMBL:Y00234 comes from this gene; cDNA EST EMBL:Y01206
comes from this gene; cDNA EST yk530c7.3 comes from this
gene"
/codon_start=1
/protein_id="CAB05544.1"
/db_xref="GI:3924820"
/db_xref="SPTREMBL:Q9XU84"
/translation="MLFPAPFISILLISFFLSLLSVVTKYQVCHGDGVNLNQHTCRPFT
EVIDAGSTGRHLHYKIDHPAIAASHGMPKVEKIFQEVKPGLSFSAKSPSAADSL
IPLOARKEVPEHFMKEKPTITLKATAGLRLPGMADDILSEVEERIFNSGFFAAPP
DAVNMPGSDGVYSWFTLILLETFLTDEPTVGHKPAAHRSVAAPFDGGGQTLATY
PNNEAVSEHGVVERDIDFRGHIRLFTSHFLGSLIAARLNILOLETDNEIESTHQL
ITSCMBEGQLTWEYALKFPWNINGSSSHSCYCYGTTKNFVSESSEIMHLRELKSPV
YLSYFLDRALNSGLVNGEGKIELQKFAAEIACRREKTEIDDGSHWMPWOCLDL
TIYSYLRDGYQFEDNOPLVLAARKIKMEASVWGGLAFANEFQLTEGAITALSSE
PNSTVVDQIFDLVYSGTNQVLSYFNISV"
24915..26580
/genes="K08H10.3"
join(24915..24953,26257..26580)
/genes="K08H10.3"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB05542.1"
/db_xref="GI:3924818"
/db_xref="SPTREMBL:Q9XU85"
/translation="MSRSPSPPTITIQQDEMSESADHLDELKSLVDEVIPVAQGV
GNQHFDEPKSQKTPPSLNSDYAEELLEAKLINTDESSVSPFLFYETPTFPNGQFPGH
PPFIFFPFLRLLLSN"
33555..36085
/genes="K08H10.1"
join(33555..33650,33695..34477,34526..35092,35188..35526,
35669..36085)
/genes="K08H10.1"
/notes="cDNA EST EMBL:D26945 comes from this gene; cDNA EST
EMBL:D26944 comes from this gene; cDNA EST EMBL:D26942 comes from this
gene; cDNA EST EMBL:D35244 comes from this gene; cDNA EST
EMBL:D32697 comes from this gene; cDNA EST EMBL:D35410
comes from this gene; cDNA EST EMBL:D35491 comes from this
gene; cDNA EST EMBL:D32827 comes from this gene; cDNA EST
EMBL:D32868 comes from this gene; cDNA EST EMBL:D68023
comes from this gene; cDNA EST EMBL:D32932 comes from this
gene; cDNA EST EMBL:D35645 comes from this gene; cDNA EST
EMBL:D35710 comes from this gene; cDNA EST EMBL:D34298
comes from this gene; cDNA EST EMBL:C10845 comes from this
gene; cDNA EST EMBL:C11544 comes from this gene; cDNA EST
EMBL:C13814 comes from this gene; cDNA EST yk194a1.3 comes
from this gene; cDNA EST yk194a1.5 comes from this gene;
cDNA EST yk213e5.3 comes from this gene; cDNA EST
yk213e5.5 comes from this gene; cDNA EST yk223c6.3 comes
from this gene; cDNA EST yk223c6.5 comes from this gene;
cDNA EST yk227f11.3 comes from this gene; cDNA EST
yk239d6.3 comes from this gene; cDNA EST yk239d6.5 comes
from this gene; cDNA EST yk239h7.3 comes from this gene;
cDNA EST yk239h7.5 comes from this gene; cDNA EST
yk246b8.5 comes from this gene; cDNA EST yk248e1.3 comes
from this gene; cDNA EST yk248e1.5 comes from this gene;
cDNA EST yk268b6.3 comes from this gene; cDNA EST
yk268b6.5 comes from this gene; cDNA EST yk273b1.3 comes
from this gene; cDNA EST yk273b1.5 comes from this gene;

```

Query Match 4.08; Score 47; DB 34; Length 39339;

Best Local Similarity 49.8%; Pred. No. 0.18;

Matches 147; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

Qy 445 cagttgtgtggcactagtgaagagcagcagcagcagcctctatgacaagtgtat 504

||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 297 CAGATCTCTGGTGGAAATGTGTCATCTCGATGATTCATCTATTCATCTCTAC 238

Qy 505 gagaacttggaggagggtggatgctgtggacaatggatctccatggcagagggg 564

Search completed: May 24, 2000, 00:38:17
Job time: 13048 sec

[illegible]

RESULT	15
CNS01B5Z	CNS01B5Z
Locus	720 bp
DEFINITION	mrna Botrytis cinerea strain T4 cdna library under conditions of nitrogen deprivation.
ACCESSION	AL113823
VERSION	AL113823.1 GI:5828442
KEYWORDS	cdna library; nitrogen deprivation.
SOURCE	Botryotinia fuckeliana.
ORGANISM	Botryotinia fuckeliana Eukaryota; Fungi; Ascomycota; Euascomycetes; Discomycetes; Leotiaceae; Sclerotiniaceae; Botryotinia. 1 (bases 1 to 720)
REFERENCE	Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y. Direct Submission
AUTHORS	Submitted (01-SEP-1999), Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
TITLE	2 (bases 1 to 720)
JOURNAL	Genoscope.
REFERENCE	Direct Submission
AUTHORS	Submitted (01-SEP-1999) Genoscope - Centre National de Sequençage ; Cp 5706 91057 ERY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
TITLE	- web : www.genoscope.cns.fr)
JOURNAL	The cdna library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.
COMMENT	

```

FEATURES      Location/Qualifiers
source        1..720
              /organism="Botryotinia fuckeliana"
              /strain="T4"
              /db_xref="taxon:40559"
              /note="Genoscope sequence ID : W675011"
base count    191 a 151 c 196 g 182 t

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:00:25 ; Search time 282.38 Seconds
(without alignments)
1039.292 Million cell updates/sec

Title: US-09-215-435-117

Perfect score: 1173

Sequence: 1 gagctgttatggacaccg.....tccatctcaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	12.4	182	1 X41222	Human secreted pro
c	35.4	3.0	1146	1 Q83793	S. avermitilis BCK
c	3	35.4	3.0	2728	1 S. avermitilis bkd
c	4	34.6	2.9	3181	1 Nucleotide sequenc
c	5	34	2.9	114955	1 Human adenosine A1
c	6	33.8	2.9	8438	1 DNA encoding pseud
7	33.6	2.9	3243	1 Q79401	Human NMDAR2 recep
8	33.6	2.9	3243	1 V82910	Human N-methyl-D-a
9	33.6	2.9	3698	1 Q79400	Human NMDAR2 recep
10	33.6	2.9	3698	1 V82909	Human N-methyl-D-a
11	33.6	2.9	4002	1 V82915	Human N-methyl-D-a
12	33.6	2.9	4017	1 V82913	Human N-methyl-D-a
13	33.6	2.9	4053	1 V82912	Human N-methyl-D-a
14	33.6	2.9	4068	1 V82914	Human N-methyl-D-a
15	33.6	2.9	4077	1 V82911	Human N-methyl-D-a
16	33.6	2.9	4092	1 Q79407	Human NMDAR2 recep
17	33.6	2.9	4274	1 Q79407	Human NMDAR2 recep
18	33.6	2.9	4289	1 Q79405	Human NMDAR2 recep
19	33.6	2.9	4325	1 Q79404	Human NMDAR2 recep
20	33.6	2.9	4340	1 Q79372	Human N-methyl-D-a
21	33.6	2.9	4349	1 Q79406	Human NMDAR2 recep
22	33.6	2.9	4364	1 Q79403	Human NMDAR2 recep
23	33	2.8	3349	1 T16518	Collagen A1/decori
24	33	2.8	3535	1 T16515	Collagen A1/BMP-2B
25	33	2.8	3541	1 T16516	Collagen A1/TGF-be
26	33	2.8	4192	1 T16517	Collagen A1/decori
27	32.8	2.8	452	1 Q28007	Sequence encoding
28	32.8	2.8	452	1 Q28760	Sequence of clone
29	32.8	2.8	53789	1 V21187	Amycolatopsis medi
30	32.6	2.8	1051	1 V73487	Human SRE-ZBP anal
31	32.6	2.8	2168	1 V73485	Human SRE-ZBP anal
32	32.6	2.8	110000	1 V30458_2	Continuation (3 of
33	32.6	2.8	110000	1 V30458_3	Continuation (4 of
34	32.6	2.8	110000	1 V30459_2	Continuation (3 of

35 32.6 2.8 110000 1 V30459_3 Continuation (4 of

36 32.4 2.8 744 1 T74073 Myceliophthora the

37 32.4 2.8 1398 1 T74027 Thermus aquaticus

c 38 32.4 2.8 1791 1 V56021 Haloperoxidase, CP

39 32.4 2.8 6232 1 Q29269 Human calcium chan

40 32.4 2.8 7175 1 Q84658 Human neuronal cal

41 32.4 2.8 7175 1 V42686 DNA encoding human

42 32.4 2.8 7266 1 V29059 Human calcium chan

43 32.4 2.8 7362 1 Q37817 Sequence encoding

44 32.4 2.8 7362 1 Q84657 Human neuronal cal

45 32.4 2.8 7362 1 V42865 DNA encoding human

ALIGNMENTS

RESULT 1

X41222
ID X41222 standard; cDNA; 182 BP.
AC X41222;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5, EST SEQ ID NO:166.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GSET) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153778/13.
DR P-PSDB; Y12389.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 385-386; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 182 BP; 26 A; 46 G; 32 T;

Query Match 12.4%; Score 146; DB 1; Length 182;
Best Local Similarity 100.0%; Pred. No. 5.8e-31;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gagctgttatggacaccgcttctctgcgcgcctcttaacgctgctgctgcgcgcgcca 60

Db 36 GAGCTGCTTATGGACACCGCTTCTCTGCGGGCTCTTACGCTGCTGCTGCGCGCCCA 95

QY 61 cccctgtatacccgccaccgcatgctcggtccagagtcctccgcccccaaacgatcc 120

Db 96 CCCCTGTATACCGGCACCGCATGCTCGGTCCAGAGTCCGCGCCGCCCAAAAGATCC 155

Wed May 24 09:19:44 2000

us-09-215-435-117.rng

QY 121 cgcagaaactaatgacccccc 146
 Db 156 CGCAGAAACTCATGCCACCCCGG 181

RESULT 2

Q83793/c
 ID Q83793 standard; cDNA: 1146 BP.
 AC Q83793;
 DT 05-SEP-1995 (first entry)
 DE S. avermitilis BCKDH El-alpha subunit.
 KW Branched-chain alpha-keto-acid-dehydrogenase; BCKDH; bkd gene;
 KW avermectin; antibiotic; acaricide; anthelmintic; insecticide;
 KW nematocide; pesticide; ds.
 OS Streptomyces avermitilis.
 FH Key Location/Qualifiers
 FT 1. 1146
 FT /*tag= a

PN WO9504150-A.
 PD 09-FEB-1995.
 PF 30-MAY-1994; IB0127.
 PR 30-JUL-1993; US-100518.
 PA (PFIZ) PFIZER INC.
 PI Denova CD;
 DR WPI: 95-082233/11.
 DR P-PSDB: R69623.
 DR Streptomyces branched-chain alpha-keto-acid dehydrogenase DNA -
 PT used to improve prodn. of natural avermectin cpds and to produce
 PT novel avermectin cpds
 PS Claim 5; Page 42; 65pp; English.
 CC Novel branched-chain alpha-keto-acid-dehydrogenase (BCKDH) genes
 CC (bkd) from Streptomyces avermitilis were cloned using PCR and
 CC homology probing. DNAs encoding the BCKDH El-alpha, El-beta, and
 CC N-terminal and internal portions of the E2 subunit (Q83793-96)
 CC were obtained. A genomic sequence including the El-alpha, El-beta
 CC and E2 (partial) bkd ORFs was also isolated (Q83797). Manipulation
 CC of bkd genes allows the enhanced production of natural or novel
 CC avermectins.
 SQ Sequence 1146 BP; 163 A; 465 C; 392 G; 126 T;

Query Match 3.0%; Score 35.4; DB 1; Length 1146;
 Best Local Similarity 49.7%; Pred. No. 1.9;
 Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 949 caaagccgctgcccctgcagagccatgcccgtcttcgagcagggccctggaccag 1008
 Db 1095 CAACATGGGGTCTCTCTCCCGCAGCTGGGGGTGGGCTCGGCATACATGCGGCGACAG 1036
 QY 1009 gtcagtggaatccctggctgcatcttcgtccatgcaagcggcttcattggcggtcacgc 1068
 Db 1035 GTCCATGGGGTCCAGGCGCGGATCTGGTTTCATGCTGCGCGCAGGTCCGCGGCATCGC 976
 QY 1069 acccgagaggggtccttgagcatgcccgtgcccacttgcccagcgctacatccca 1128
 Db 975 CTCGGGCTCTCGCGGCGCGGATGCCGTCTCGTGCAGCAGCCCGGTTGGTCAA 916
 QY 1129 c 1129
 Db 915 c 915

RESULT 3

Q83797/c
 ID Q83797 standard; cDNA: 2728 BP.
 AC Q83797;
 DT 05-SEP-1995 (first entry)
 DE S. avermitilis bkd gene region.
 KW Branched-chain alpha-keto-acid-dehydrogenase; BCKDH; bkd gene;
 KW avermectin; antibiotic; acaricide; anthelmintic; insecticide;
 KW nematocide; pesticide; ds.
 OS Streptomyces avermitilis.
 FH Key Location/Qualifiers

FT rbs 390. 395
 FT /*tag= a
 FT 403. 1548
 FT /*tag= b
 FT /product= El-alpha subunit
 FT 1607. 1614
 FT /*tag= c
 FT 1622. 2626
 FT /*tag= d
 FT /product= El-beta subunit
 FT 2610. 2615
 FT /*tag= e
 FT 2626. 2728
 FT /*tag= f
 FT /product= E2 N-terminal sequence

WO9504150-A.
 PN 09-FEB-1995.
 PD 30-MAY-1994; IB0127.
 PF 30-JUL-1993; US-100518.
 PR (PFIZ) PFIZER INC.
 PA Denova CD;
 PI WPI: 95-082233/11.
 DR P-PSDB: R69623; R69624; R69625.
 DR Streptomyces branched-chain alpha-keto-acid dehydrogenase DNA -
 PT used to improve prodn. of natural avermectin cpds and to produce
 PT novel avermectin cpds
 PS Claim 5; Page 44-45; 65pp; English.
 CC Novel branched-chain alpha-keto-acid-dehydrogenase (BCKDH) genes
 CC (bkd) from Streptomyces avermitilis were cloned using PCR and
 CC homology probing. DNAs encoding the BCKDH El-alpha, El-beta, and
 CC N-terminal and internal portions of the E2 subunit (Q83793-96)
 CC were obtained. A genomic sequence including the El-alpha, El-beta
 CC and E2 (partial) bkd ORFs was also isolated (Q83797). Manipulation
 CC of bkd genes allows the enhanced production of natural or novel
 CC avermectins.
 SQ Sequence 2728 BP; 388 A; 1005 C; 963 G; 372 T;

Query Match 3.0%; Score 35.4; DB 1; Length 2728;
 Best Local Similarity 49.7%; Pred. No. 2.6;
 Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 949 caaagccgctgcccctgcagagccatgcccgtcttcgagcagggccctggaccag 1008
 Db 1497 CAACATGGGGTCTCTCTCCCGCAGCTGGGGGTGGGCTCGGCATACATGCGGCGACAG 1438
 QY 1009 gtcagtggaatccctggctgcatcttcgtccatgcaagcggcttcattggcggtcacgc 1068
 Db 1437 GTCCATGGGGTCCAGGCGCGGATCTGGTTTCATGCTGCGCGCAGGTCCGCGGCATCGC 1378
 QY 1069 acccgagaggggtccttgagcatgcccgtgcccacttgcccagcgctacatccca 1128
 Db 1377 CTCGGGCTCTCGCGGCGCGGATGCCGTCTCGTGCAGCAGCCCGGTTGGTCAA 1318
 QY 1129 c 1129
 Db 1317 c 1317

RESULT 4

V59358
 ID V59358 standard; cDNA: 3181 BP.
 AC V59358;
 DT 14-DEC-1998 (first entry)
 DE Nucleotide sequence encoding human type 1 (alpha-1) collagen.
 KW ss: human; type 1 (alpha-1) collagen; trans-4-hydroxyproline;
 KW 3-hydroxyproline; post-translational hydroxylation; collagen.
 OS Homo sapiens.
 PN US5821089-A.
 PD 13-OCT-1998.
 PF 03-JUN-1996; 655086.
 PR 03-JUN-1996; US-655086.
 PA (BUEC/) BUECHTER D. D.

CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC metastases, as well as all types of cancers which may metastasize or have
CC metastasized to the lungs, including breast and prostate cancer.
SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

	Query Match	2.9%; Score 34; DB 1; Length 114955;
	Best Local Similarity	37.1%; Pred. No. 25;
	Matches 92; Conservative	12; Mismatches 144; Indels 0; Gaps 0;

```

QY      19 cgcttcctgcgcgcccttaacgtactgtcgtccgcccaaaacgatccccgcagcaaatcatggcac 78
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    108590 CGCGCGCGCCCCGCCCNHNHNSCAGGCGCGCGCGCGCGCCGCCCNHNHNSGCCAGGCGC 108531

QY      79 cgcattgctcgttcagagtccgtccgcccccacgaatccccgcagcaaatcatggcac 138
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    108530 CGCGCGCGCGCCCNHNHNSAGCAGGCGCGCGCGCGCCGCCCNHNHNSGAGCCAGCGC 108471

QY      139 ccgccccgaatcgggagcgacaatggcaaccttccactgcagcgaggcaactggcatgcgca 198
          ||| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    108470 CCGCGCGCGCGCCCNHNHNSCGAGCCAGGCGCGCGCGCGCCGCCCNHNHNSGGCGAGCC 108411

QY      199 ctgcttcgctcttcgcggagtaccgggatcacagattgtcgaccgccggatcccgcaa 258
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    108410 AGCGCGCGCGCGCGCGCCGCCCNHNHNSGGCGAGCCAGGCGCGCGCGCGCCGCCCNHNHNS 108351

QY      259 aaactcgc 266
          | | | | |
Db    108350 AGGCGAGC 108343

```

RESULT 6

ID	Q73500/c	
AC	Q73500 standard; DNA; 8438 BP.	
DC	15-MAY-1995 (first entry)	
DE	DNA encoding Pseudorabies virus large latency transcript.	
KW	Pseudorabies virus; PRV; LLV; large latency transcript;	
KW	attenuated virus; vaccine; early protein 0; EP0; HSV-1 ICPO;	
KW	protecting animals; deletion mutants; swine; ds.	
OS	Pseudorabies virus.	
FH	Key Location/Qualifiers	
FT	misc_feature	1..7013
FT		/tag= a
FT		/note= "derived from PRV strain InFh"
FT	misc_feature	7014..8425
FT		/tag= b
FT		/note= "derived from PRV strain Ka"
FT	cds	622..6498
FT		/tag= c
FT	tata_signal	1..6
FT		/tag= d
FT	misc_feature	34
FT		/tag= e
FT		/note= "RNA cap site"
FT	polya_signal	8382..8387
FT		/tag= f
PN	US5352596-A.	
PD	04-OCT-1994.	
PP	11-SEP-1992; 945283.	
PR	11-SEP-1992; US-945283.	
PA	(USDA) US SEC OF AGRIC.	
PI	Cheung AK, Wesley RD;	
DR	WPI; 94-316187/39.	
DR	P-PSDB; R60620.	
PT	New pseudorabies virus mutants for use in vaccine - having a	
PT	deletion and/or insertion in the early protein O gene or large	

us-09-215-435-117.rng

Wed May 24 09:19:44 2000

CC adds an extra 3 a.a. to the peptide sequence but causes a reading frame
 CC shift which terminates the peptide at residue 494 of the NMDAR2C
 CC sequence. The NMDA receptor contains two subunits: subunit R1 (Q79370)
 CC and subunit R2 selected from the subunits 2A (Q79375), 2B (Q79377), 2C
 CC and 2D (Q79378). The receptor forms part of a family of NMDA receptors
 CC which have cation-selective channels and bind glutamate and NMDA. The
 CC NMDAR2C gene was obtained by amplifying cDNA derived from human brain
 CC tissues with primers corresponding to sequences in the rat NMDAR2A
 CC receptor DNA and using the resultant fragments as probes to screen a cDNA
 CC library derived from human hippocampal RNA. 4 basic clones were isolated:
 CC NMDA21 (Q79399), NMDA22 (Q79400), NMDA24 (Q79401) and NMDA26 (Q79402).
 CC The clones are thought to be splice variants of each other. Based on the
 CC sequence of the 4 clones, a series of variants (Q79403-7) of the NMDAR2C
 CC receptor were constructed. The expression of the genes allows the
 CC reconstruction of the NMDA receptor. The complete receptor can be used
 CC to identify compounds which bind or are antagonistic to the human NMDA
 CC receptor. 3243 BP; 540 A; 1115 C; 1024 G; 564 T;
 SQ Sequence 3243 BP; 540 A; 1115 C; 1024 G; 564 T;
 Query Match 2.9%; Score 33.6; DB 1; Length 3243;
 Best Local Similarity 52.9%; Pred. No. 8.7;
 Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 Qy 157 cacaatggcaccctccactgcgacgagcagcactgcatgacgactgcttcgctccctgccc 216
 Db 2445 CACTACAGCTCTTCTCTCGAGCGACGATCCGCGCCCTCTCTCCGCTCTTCCCG 2504
 Qy 217 gactacgggagtcagagattgtgcgacccgggacccgaaactcgttctctgtgac 276
 Db 2505 GAGCCCCCGGAGCTGAGGACCTCCGCTGCTCGGAGCAGCTGGCCCGCGGGAG 2564
 Qy 277 atcgtgtggacgtgg 292
 Db 2565 GCCCTGCTGCACGCGG 2580
 RESULT 8
 V82910
 ID V82910 standard; cDNA; 3243 BP.
 AC V82910;
 DT 23-FEB-1999 (first entry)
 DE Human N-methyl-D-aspartate receptor subunit clone NMDA24.
 KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
 KW NMDA-activated cation-selective ion channel; glutamate receptor; ss.
 OS Homo sapiens.
 FH Key
 FT CDS
 FT 1..3243
 FT /tag= a
 FT /product= NMDA_receptor_subunit
 FT /note= "no termination codon"
 FT US5849895-A.
 PN 15-DEC-1998.
 PD 20-APR-1994; 231193.
 PF 20-APR-1994; US-231193.
 PR 20-APR-1993; US-052449.
 PR (SIBI-) SIBIA NEUROSCIENCES INC.
 PA Daggett LP, Lu C;
 PI WPI; 99-069812/06.
 DR P-PSDB; W87504.
 DR DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
 PT assembly of functional glutamate receptor subunits
 PS example 3; Columns 261-268; 203pp; English.
 PS The present sequence encodes a human N-methyl-D-aspartate (NMDA)
 CC receptor subunit (NMDAR). The nucleic acid sequence does not contain the
 CC 860 5'-most nucleotides, has an additional 11 nucleotides (V82891)
 CC inserted between nucleotides 1300 and 1301, an additional 24 nucleotides (V82890)
 CC cDNA sequence is derived from clone NMDA24. The NMDAR subunits contribute
 CC to the formation of NMDA-activated cation-selective ion channels. In
 CC addition to being useful for the production of NMDA receptor subunit
 CC proteins, the nucleic acids are also useful as probes to identify and
 CC isolate nucleic acids encoding related receptor subunits. Functional

PT latency transcript gene
 PS Disclosure; Column 15-30; 43pp; English.
 CC Q73500 shows the pseudorabies virus (PRV) large latency transcript
 CC (LRT). The basic sequence is derived from PRV strain InPh and PRV
 CC strain Ka. The LRT overlaps and is transcribed in the opposite
 CC orientation with respect to the EP0 (early polypeptide 0) and the
 CC immediately early gene (IE180). EP0 is nonessential for replicatio,
 CC LRT is the only gene expressed during PRV latency, and the IE180
 CC gene is absolutely necessary for PRV replication. However there are
 CC 2 copies of IE180 in the genome. It is expected that PRV lacking one
 CC of the IE180 copies is viable. Deletions in the non-overlapping
 CC regions of these 3 genes will generate single deletion routants,
 CC while deletions in overlapping regions will generate double deletion
 CC mutants. The invention is concerned with the construction of attenuated
 CC viruses which have a reduced ability to reactivate from latency. This
 CC can be achieved by functionally disabling the expression of the EP0
 CC gene, or by disrupting the synthesis of the LRT, or both. (See also
 CC Q73501 and R60620-24)
 SQ Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T;
 Query Match 2.9%; Score 33.8; DB 1; Length 8438;
 Best Local Similarity 51.7%; Pred. No. 11;
 Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 Qy 965 tqccagagccatgcccgggtcttcggagcagccctggaccaggtgagtgatccctg 1024
 Db 4562 TCCCGGGGGCGGGCGGCGGAGCGGAGCAGCGCGGAGCTCTGTGCGCTCTGTCGCC 4503
 Qy 1025 gtgcgatcttgcgcagcagggcttcattgctgctcaccgacccagaggggtgcct 1084
 Db 4502 GCACCGTGGCGCGCTGTGTGCTGTACAGCTGAGCGGGCGCGCGCGCGCGCGCT 4443
 Qy 1085 tqagcatgcccgtgcacccttggcccg 1113
 Db 4442 GGACCTACCGCGCGCGCTCTTCGCCCG 4414
 RESULT 7
 Q79401
 ID Q79401 standard; cDNA; 3243 BP.
 AC Q79401;
 DT 06-JUL-1995 (first entry)
 DE Human NMDAR2 receptor subunit clone NMDA24 gene.
 KW N-methyl-D-aspartate; receptor; human; NMDA; cation-selective; channel;
 KW glutamate; hippocampus; rat; pcDNA1; NMDA receptor; antagonist; ds.
 OS Homo sapiens.
 FH Key
 FT CDS
 FT 1..822
 FT /tag= a
 FT /product= part of the NMDAR2C receptor subunit
 FT /note= "11 bp insertion found in clone NMDA24"
 FT /tag= b
 FT /note= "1525
 FT /tag= c
 FT /note= "24 bp insertion found in clone NMDA24"
 FT W09424284-A.
 PN 27-OCT-1994.
 PD 20-APR-1994; US04387.
 PF 20-APR-1993; US-052449.
 PR (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PA Daggett LP, Ellis SB, Liaw CW, Lu C;
 PI WPI; 94-341863/42.
 DR P-PSDB; R66059.
 DR Isolated DNA encoding a human N-methyl-D-aspartate receptor
 PT subunit - used as probes in the identification and isolation of
 PT nucleic acids encoding related receptor subunits.
 PS Claim 11; Page 99-104; 156pp; English.
 CC The nucleotide sequence of the novel N-methyl-D-aspartate (NMDA) receptor
 CC 2C (NMDAR2C) gene clone NMDA24. This clone covers bases 861-4068 of the
 CC complete NMDAR2C gene (Q79372) with an insertion of 11 bp between bases
 CC 1300-1 and an insertion of 24 bp between bases 2350-1. The 11 bp insert

2

CC to identify and isolate nucleic acids encoding related receptor subunits.
CC Functional glutamate receptors can be assembled from several NMDA
CC receptor subunit proteins of one type (homomeric) or from combinations
CC of subunit proteins of different types (heteromeric). The present
CC invention also comprises methods for using such receptor subunits to
CC identify and characterise compounds which affect the function of such
CC receptors, e.g. agonists, antagonists and modulators of glutamate
CC receptor function. The invention also comprises methods for determining
CC whether unknown protein(s) are functional as NMDA receptor subunits.
SQ Sequence 4077 BP; 664 A; 1421 C; 1271 G; 720 T;

Query Match 2.9%; Score 33.6; DB 1; Length 4077;
Best Local Similarity 52.9%; Pred. No. 9.5;
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 157 cacaatggacaccttcactgcagcagggcactgcatgcatgctctcctcctgcg 216
DB 3279 CACTACAGCTCCTTTCTTCGAGCGGACCGATCCGGCGCCCTTCTCCGGCTCTTCCCG 3338
QY 217 gactaccgggatgcagagattgtgcggaccgccgggattcccgaaaaactcgtctctgtgac 276
DB 3339 GAGCCCCCGGAGCTGGAGGACCTGCCGCTGCTCGGTCCGGAGCAGCTGGCCCGGGGAG 3398
QY 277 atcgtgtggacgtgg 292
DB 3399 GCCCTGCTGAACGCG 3414

Search completed: May 24, 2000, 00:02:10
Job time: 10816 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:55:11 ; Search time 203.6 seconds
(without alignments)
748.882 Million cell updates/sec

Title: US-09-215-435-117

Perfect score: 1173

Sequence: 1 gagctgtctatggacacgcg.....tccatctcaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

- 1: /cgn2.6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2.6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2.6/ptodata/1/ina/5C_COMB.seq:*
- 4: /cgn2.6/ptodata/1/ina/5D_COMB.seq:*
- 5: /cgn2.6/ptodata/1/ina/6_COMB.seq:*
- 6: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq:*
- 7: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1163.4	99.2	1380	4	US-08-933-750C-74	Sequence 74, Appl
C 2	36.6	3.1	1218	1	US-08-232-463-14	Sequence 14, Appl
C 3	35.4	3.0	1146	2	US-08-482-385A-1	Sequence 1, Appl
C 4	35.4	3.0	2728	2	US-08-482-385A-5	Sequence 5, Appl
C 5	34.6	2.9	3181	2	US-08-655-086-1	Sequence 1, Appl
C 6	33.8	2.9	8438	1	US-07-945-283-1	Sequence 1, Appl
C 7	33.6	2.9	3243	3	US-08-231-193A-44	Sequence 44, Appl
C 8	33.6	2.9	3243	4	US-08-486-273A-44	Sequence 44, Appl
C 9	33.6	2.9	3243	5	US-08-480-474-44	Sequence 44, Appl
C 10	33.6	2.9	3698	3	US-08-231-193A-43	Sequence 43, Appl
C 11	33.6	2.9	3698	4	US-08-486-273A-43	Sequence 43, Appl
C 12	33.6	2.9	3698	5	US-08-480-474-43	Sequence 43, Appl
C 13	33.6	2.9	4002	3	US-08-231-193A-53	Sequence 53, Appl
C 14	33.6	2.9	4002	4	US-08-486-273A-53	Sequence 53, Appl
C 15	33.6	2.9	4002	5	US-08-480-474-53	Sequence 53, Appl
C 16	33.6	2.9	4017	3	US-08-231-193A-49	Sequence 49, Appl
C 17	33.6	2.9	4017	4	US-08-486-273A-49	Sequence 49, Appl
C 18	33.6	2.9	4017	5	US-08-480-474-49	Sequence 49, Appl
C 19	33.6	2.9	4053	3	US-08-231-193A-47	Sequence 47, Appl
C 20	33.6	2.9	4053	4	US-08-486-273A-47	Sequence 47, Appl
C 21	33.6	2.9	4053	5	US-08-480-474-47	Sequence 47, Appl
C 22	33.6	2.9	4058	3	US-08-231-193A-5	Sequence 5, Appl
C 23	33.6	2.9	4058	4	US-08-486-273A-5	Sequence 5, Appl
C 24	33.6	2.9	4068	5	US-08-480-474-5	Sequence 5, Appl
C 25	33.6	2.9	4077	3	US-08-231-193A-51	Sequence 51, Appl
C 26	33.6	2.9	4077	4	US-08-486-273A-51	Sequence 51, Appl
C 27	33.6	2.9	4077	5	US-08-480-474-51	Sequence 51, Appl

Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 7, Appl

28 33.6 2.9 4092 3 US-08-231-193A-45
29 33.6 2.9 4092 4 US-08-486-273A-45
30 33.6 2.9 4092 5 US-08-480-474-45
31 32.8 2.8 452 1 US-07-662-198B-1
32 32.8 2.8 452 1 US-08-322-742-1
33 32.8 2.8 1394 4 US-08-068-729-3
34 32.8 2.8 1950 5 US-08-685-466C-1
35 32.4 2.8 5467 1 US-07-745-206A-12
36 32.4 2.8 5467 3 US-08-311-363-12
37 32.4 2.8 7175 2 US-08-455-543A-8
38 32.4 2.8 7175 3 US-08-193-078B-8
39 32.4 2.8 7175 3 US-08-223-305C-8
40 32.4 2.8 7175 3 US-08-149-097D-8
41 32.4 2.8 7266 5 US-08-713-118-1
42 32.4 2.8 7362 2 US-08-455-543A-7
43 32.4 2.8 7362 3 US-08-193-078B-7
44 32.4 2.8 7362 3 US-08-223-305C-7
45 32.4 2.8 7362 3 US-08-149-097D-7

ALIGNMENTS

RESULT 1
US-08-933-750C-74
; Sequence 74, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/08/933,750C
; APPLICATION NUMBER: 536
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNO00703

CLONE: 1572888
US-08-933-750C-74

Query Match 99.2%; Score 1163.4; DB 4; Length 1380;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gagctgttatggagacacgcttctctgagcgccctcttaacgctgctgctgcccgcgca 60
DB 46 GAGCTGCTTATGGACACGCTTCTGCGCGCCCTCTTAAACGCTGCTGCTGCCGCGCA 105
QY 61 ccctgtataccggcaccgcgcatgctgctccagagtcgctcccccacaaacgatcc 120
DB 106 CCCTGTATACCGGCACCGCATGCTCGTCCAGAGTCGTCCTCCGCCCCCAACGATCC 165
QY 121 cgcagcaaatcatgacacgcccgcgaatcgggagcagcaaatggcaccttcactgcgac 180
DB 166 CGCAGCAAACTCATGACGACCGCCCGCAATCGGGAGCGCAATGGCACCTTCCACTGCCAC 225
QY 181 gaggcactgcgacgacactgcttcctcctcctgcgagtgacccgggatgcagattgtg 240
DB 226 GAGCAGCTGCATGCCACTGCTTCGCTCCTCGCGAGTACCGGATCGCGGATGCAGAGATTGTG 285
QY 241 cggaccgggatcccgaaacactcgttctcgtgacatcgtggtgacgtggggcgaag 300
DB 286 CGGACCGGGATCCGAAAACTCGCTTCTGTGACATCGTGGTGACGCTGGGGGCGAG 345
QY 301 taagacctcgagacacccgatgatgaccatcaccagaggtctttcacagagaccatgagc 360
DB 346 TACGACCTCGGAGACACCGCATATGACCATCACAGAGTCTTTCACAGAGACCATGAGC 405
QY 361 tccctgtccctggagcgctggcagacaaagtgcagagtgctggggaactatctatctg 420
DB 406 TCCCTGTCCCTGGAAAGCGCTGCAGACCAAGCTGAGCAGTGGGAGCTCATCTATCTG 465
QY 421 cacttgggcaacagctgctgcccagtttctgagcactagtgaaagagacagcatgggtg 480
DB 466 CACTTCGGGCAACAGCTGCTGGCCAGTGTCTGGGCACTAGTGAAGAGGACACATGGTG 525
QY 481 ggcacctctatgacaaagtatgagaactttgtggagagtggtggtggtggaacat 540
DB 526 GGCACCTCTATGACAGATGATGAGAACTTTGTGGAGGAGTGGATGCTGTGGACAAT 585
QY 541 gggatctccagtgggcagagggagcctcgatgatacactgaccactaccctgagtgca 600
DB 586 GGGATCTCCAGTGGGAGAGGGGAGGCTCGATGACCTGACCTGACCATACCTCCCTGAGTGA 645
QY 601 csagtgctcgacttaactctacctggaaaccccccacacacacacacagcaggggtc 660
DB 646 CGAGTTGCTCGACTTAATCTTACCTGGAACCCACCCGACCAAGACACTGAGGAGGGTTC 705
QY 661 aagcgtgcaatggatctggttcaagagagttctgcagagattagattttctaccaacac 720
DB 706 AAGCGTGCAATGGATCTGGTTCAAGAGAGATTCTGCGAGAGATTAGATTCTTACCAACAC 765
QY 721 agctgggtccagccgggctgtgaaagggcccttgcagagccttgcagcagattccagtgagc 780
DB 766 AGCTGGCTGCCAGCCGGCCCTTGTGGNAGAGGCCCTTCCAGCGGATTCAGGTGGAC 825
QY 781 ccaagtggagagattgtgaaactgagcgaagtgcatgtccctggaaagagacatctctac 840
DB 826 CCAAGTGGAGAGATTGTGAACTGGGAAAGGTGATGTCTCCCTGGAAGGAGCATCTCTAC 885
QY 841 cacttggatctgggtgtccctccagtgccatctcttctgttatcactgaacag 900
DB 886 CACTTGAATCTGGGTGTGCCCTCCAGTGGCACTCTTCTTTGTTATCTACACTGACCAG 945
QY 901 gctggacagtggcgaatacagtggtgcccgaaggagcccaactcatctccaaagccgggtg 960
DB 946 CTTGGACAGTGGGGAATACAGTGTGTGCCAAGGAGGCCCACTCATTTCCAAAGCCGGCTG 1005
QY 961 ccctcgccagagccatgagcggggtcttcgggagcagggccctgagcaggtcagtgggatc 1020

RESULT 2

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; -US-08-232-463-14

Query Match 3.1%; Score 36.6; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 0.86;
Matches 15; Conservative 208; Mismatches 172; Indels 0; Gaps 0;

QY 459 tagtgaagagacagcatgtggcaccctctatgacaagatgatgagaactttgtgga 518
DB 1440 TGGTACRR 1381


```

ADDRESS:  Curtis P. Ribando
STREET:  1815 No. 5352596th University Street
CITY:    Peoria
STATE:   IL
COUNTRY: USA
ZIP:     61604

COMPUTER READABLE FORM:
MEDIUM TYPE:  Floppy disk
COMPUTER:     IBM PC compatible
OPERATING SYSTEM:  PC-DOS/MS-DOS
SOFTWARE:      PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/07/945,283
FILING DATE:       19920911
CLASSIFICATION:    424
ATTORNEY/AGENT INFORMATION:
NAME:              Ribando, Curtis P
REGISTRATION NUMBER:  27976
TELECOMMUNICATION INFORMATION:
TELEPHONE:          309-685-4011 ext.513
TELEFAX:            309-685-4128
INFORMATION FOR SEQ ID NO:  1:
SEQUENCE CHARACTERISTICS:
LENGTH:             8438 base pairs
TYPE:                NUCLEIC ACID
STRANDEDNESS:       double
TOPOLOGY:            linear
MOLECULE TYPE:      DNA (genomic)
HYPOTHETICAL:       NO
ANTI-SENSE:         NO
ORIGINAL SOURCE:
ORGANISM:            Pseudorabies virus
FEATURE:
NAME/KEY:            CDS
LOCATION:              622..6495
FEATURE:
NAME/KEY:            variation
LOCATION:              replace(1099, "g")
FEATURE:
NAME/KEY:            variation
LOCATION:              replace(1267, "t")
FEATURE:
NAME/KEY:            variation
LOCATION:              replace(1381, "c")
FEATURE:
NAME/KEY:            variation
LOCATION:              replace(1566, "c")
FEATURE:
NAME/KEY:            variation
LOCATION:              replace(7010, "g")

```

	Query Match	2.9%;	Score 33.8;	DB 1;	Length 8438;
	Best Local Similarity	51.7%;	Pred. No. 5.5;		
	Matches	77;	Conservative	0;	Mismatches
				72;	Indels
				0;	Gaps
QY	965	tgccagagccatggcggggtcttcggagcagaggcccttgaccagtcagtgaggatccctg	1024		
Db	4562	TCCCCGGGGCGGCGCCGGCCGAGGCGGACGAGCGCGGACTCTGTGCGCCCTCTGTCGCC	4503		
QY	1025	gctgcactcttgctccatgtcaagcggcttcattgctgggtcacccgaccccgagaggtgacct	1084		
Db	4502	GCACCGTGGCGCCGCTGGTGGCGCTACACGCTGGACGGGGCCCGCCCGGACGCCGCT	4443		
QY	1085	tgagcatggccctggccaccttggcccaag	1113		
Db	4442	GGACTACGCGCGGCGCCCTCTTCGCCCG	4414		

RESULT 7
US-08-231-193A-44
: Sequence 44, Application US/08231193A

QY 1129 c 1129
Db 1317 c 1317

RESULT 5
US-08-655-086-1
: Sequence 1, Application US/08655086
: Patent No. 5821089
: GENERAL INFORMATION:
: APPLICANT: GRUSKIN, ELLIOTT A.
: APPLICANT: BUECHTER, DOUGLAS
: APPLICANT: ZHANG, GUANGHUI
: APPLICANT: CONNOLLY, KEVIN
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DILWORTH & BARRESE
: STREET: 333 EARLE OVINGTON BOULEVARD
: CITY: UNIONDALE
: STATE: NY
: COUNTRY: US
: ZIP: 11553
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/655,086
: FILING DATE: 03-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: STEEN, JEFFREY S.
: REGISTRATION NUMBER: 32,063
: REFERENCE/DOCKET NUMBER: 203-1632
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 516-228-8484
: TELEFAX: 516-228-8516
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3181 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: CDNA
US-08-655-086-1

	Query Match	2.9%	Score 34.6;	DB 2:	Length 3181;
	Best Local Similarity	56.6%;	Pred. No. 2.2;	Mismatches	Indels
	Matches 64;	Conservative	49:	Gaps	0:
QY 43	ctgtgctgcggccgacccccctgtatacccggcaccgatctcggttcagatccgtc	102			
Db 3054	CTGTTGGTCCCGGCCCTCTGGACCTCCTGGTGTCCTGTGTCTCTCCAGCTGTTT	3113			
QY 103	cggcccccaaacgattccgcgagcaaatcatggcacgcccccgaatcgggac	155			
Bb 2114	TGTCATTGCAATGCCTCTCCCCAGCACCTTAAGAAGACTCACCATGTGTGC	3166			

RESULT 6
US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

RESULT 8
US-08-486-273A-44
; Sequence 44, Application US/08486273A
; Patent No. 598586
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.

RESULT 9
US-08-480-474-44
; Sequence 44, Application US/08480474
; Patent No. 6033865
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METH
; TITLE OF INVENTION: SAME AND USE

```

;
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,474
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9382B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0062
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...3243
;
; US-08-480-474-44
;
;
; Query Match 2.9%; Score 33.6; DB 5; Length 3243;
; Best Local Similarity 52.9%; Pred. No. 4.2;
; Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
;
; QY 157 cacaatggcaccctccactgcgacggagcactggcagcactgttgcctctctgcg 216
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 2445 CACTACAGCTCTTCTTCGAGCGGACCGATCGCGCGCGCTTCTCGCGTCTTCCG 2504
;
; QY 217 gactacgggatgcagagattgtcgaccggatcccgaaactcgtctctgtgac 276
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 2505 GAGCCCCCGAGGTGGAGACCTGCTCGTCCGGAGCAGCTGGCCCGCGGGAG 2564
;
; QY 277 atcgtgtgtgacgtgg 292
; ||| ||| ||| |||
; Db 2565 GCCCTGCTGAACGGG 2580
;
; RESULT 10
; US-08-231-193A-43
; Sequence 43, Application US/08231193A
; Patent No. 5849895
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,474
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9382B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0062
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...3243
;
; US-08-480-474-44
;
;
; Query Match 2.9%; Score 33.6; DB 5; Length 3243;
; Best Local Similarity 52.9%; Pred. No. 4.2;
; Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
;
; QY 157 cacaatggcaccctccactgcgacggagcactggcagcactgttgcctctctgcg 216
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 2445 CACTACAGCTCTTCTTCGAGCGGACCGATCGCGCGCGCTTCTCGCGTCTTCCG 2504
;
; QY 217 gactacgggatgcagagattgtcgaccggatcccgaaactcgtctctgtgac 276
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 2505 GAGCCCCCGAGGTGGAGACCTGCTCGTCCGGAGCAGCTGGCCCGCGGGAG 2564
;
; QY 277 atcgtgtgtgacgtgg 292
; ||| ||| ||| |||
; Db 2565 GCCCTGCTGAACGGG 2580
;
; RESULT 11
; US-08-486-273A-43
; Sequence 43, Application US/08486273A
; Patent No. 5985586
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA
; ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,193A
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,459
; FILING DATE: 20-APR-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0099
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 3...3698
;
; US-08-231-193A-43
;
;
; Query Match 2.9%; Score 33.6; DB 3; Length 3698;
; Best Local Similarity 52.9%; Pred. No. 4.5;
; Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
;
; QY 157 cacaatggcaccctccactgcgacggagcactggcagcactgttgcctctctgcg 216
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 2900 CACTACAGCTCTTCTTCGAGCGGACCGATCGCGCGCGCTTCTCGCGTCTTCCG 2959
;
; QY 217 gactacgggatgcagagattgtcgaccggatcccgaaactcgtctctgtgac 276
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 2960 GAGCCCCCGAGGTGGAGACCTGCTCGTCCGGAGCAGCTGGCCCGCGGGAG 3019
;
; QY 277 atcgtgtgtgacgtgg 292
; ||| ||| ||| |||
; Db 3020 GCCCTGCTGAACGGG 3035
;
; RESULT 11
; US-08-486-273A-43
; Sequence 43, Application US/08486273A
; Patent No. 5985586
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA
; ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,273A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 3...3698
; US-08-486-273A-43

```

	Query Match	2.9%	Score 33.6;	DB 4;	Length 3698;
	Best Local Similarity	52.9%;	Pred. No. 4.5;		
	Matches	72;	Conservative	0;	Mismatches 64; Indels 0; Gaps 0;
Qy	157	cacaatgacacttccactgcgacgaggcactgcatgctgcactgcttcgctcctgcgcg	216		
Db	2900	CACACAGCTCCTTCTCGAGCGCAGCGATCCGGCGGCCCTTCTCTCCCGCTCTTCCCG	2959		
Qy	217	gagtaccgggatgcagagattgtcggaccgcgggatccgcaaaactcgtcttcctgtgac	276		
Db	2960	GAGCCCCGGAGCTGGAGGACCTGCCGCTGTCGTCGGAGCAGCTGGCCCCGGGGGAG	3019		
Qy	277	atcgtggtgacgtgg	292		
Db	3020	GCCTGTCTGAACGGG	3035		

```

RESULT 12
US-08-480-474-43
; Sequence 43, Application US/08480474
; Patent No. 6033865
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA ENCODING
; TITLE OF INVENTION: SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,474
; FILING DATE: 06-JUN-1995

```

```

; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9382B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 3...3698
; US-08-480-474-43

Query Match          2.9%; Score 33.6; DB 5; Length 3698;
Best Local Similarity 52.9%; Pred. No. 4.5;
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 157 cacaatggcaccttcactgcagcaggcactggcactggcactgcttcgcctcctgcg 216
      ||| || |||| || |||| || |||| || |||| || |||| || |||| || ||
Db 2900 CACTACACGCTCTTCTCTCGAGCCGACGATCCGGCGGCCCTTCTCTCCGCTCTTCCCG 2959

Qy 217 gagtaccgggatgcagagatgtgcgaccgggataccgaaaaactgccttcctgtgac 276
      ||| || |||| || |||| || |||| || |||| || |||| || |||| || ||
Db 2960 GAGCCCCCGAGCTGAGGACCTGCCGCTGCTCGGTCCGGAGCAGTGGCCCGGGGAG 3019

Qy 277 atcgtggtggacgtgg 292
      ||| || |||| || ||
Db 3020 GCCCTGCTGAACGGCG 3035

RESULT 13
US-08-231-193A-53
; Sequence 53, Application US/08231193A
; Patent No. 5849895
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,193A
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,459
; FILING DATE: 20-APR-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779

```

Wed May 24 09:19:44 2000

```

; REFERENCE/DOCKET NUMBER: 6362-9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 189..3833
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 189..3833
; US-08-231-193A-53

Query Match      2.9%; Score 33.6; DB 3; Length 4002;
Best Local Similarity 52.9%; Pred. No. 4.6;
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 157 cacaatggcaccttcacatgcagagcactggcagcactgcttcgctcctgcgcg 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3204 CACTACAGCTCTTCTCTCGAGCCGACCGATCCGGCGCCCTTCTCCGCTCTTCCCG 3263

QY 217 gactacgggagtcagagattgtgcgaccgggacccgaaactcgtctcctgtgac 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3264 GAGCCCCCGAGCTGGAGGACCTGCCGCTGCTCGGTCGGGAGCAGCTGGCCCGGGAG 3323

QY 277 atcgtgtgagctgg 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3324 GCCCTGCTGAACGGG 3339

RESULT 14
US-08-486-273A-53
; Sequence 53, Application US/08486273A
; Patent No. 5985586
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA
; TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,273A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062

; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 189..3833
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 189..3833
; US-08-486-273A-53

Query Match      2.9%; Score 33.6; DB 4; Length 4002;
Best Local Similarity 52.9%; Pred. No. 4.6;
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 157 cacaatggcaccttcacatgcagagcactggcagcactgcttcgctcctgcgcg 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3204 CACTACAGCTCTTCTCTCGAGCCGACCGATCCGGCGCCCTTCTCCGCTCTTCCCG 3263

QY 217 gactacgggagtcagagattgtgcgaccgggacccgaaactcgtctcctgtgac 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3264 GAGCCCCCGAGCTGGAGGACCTGCCGCTGCTCGGTCGGGAGCAGCTGGCCCGGGAG 3323

QY 277 atcgtgtgagctgg 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3324 GCCCTGCTGAACGGG 3339

RESULT 15
US-08-480-474-53
; Sequence 53, Application US/08480474
; Patent No. 6033865
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA ENCODING
; TITLE OF INVENTION: SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,474
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9382B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:01:46 ; Search time 3550.87 Seconds
(without alignments)

-1377.221 Million cell updates/sec

Title: US-09-215-435-117

Perfect score: 1173

Sequence: 1 gagctcgcttatgggacaccg.....tccatctcaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues

Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending_Patents_NA:*

```
1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq.*
```

```
45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US095_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq.*
86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq.*
87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq.*
88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq.*
90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq.*
91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq.*
92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq.*
93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq.*
94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq.*
95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq.*
96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq.*
97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq.*
98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	IDB	Description
1	1173	100.0	1173	43	US-09-215-435-117 Sequence 117, App
2	1173	100.0	1173	72	US-60-081-563-74 Sequence 74, Appl
3	1169.8	99.7	1456	57	US-09-515-126-14066 Sequence 14066, A


```
;
; OTHER INFORMATION: identity 91
; OTHER INFORMATION: region 7..112
; OTHER INFORMATION: id H50235
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 843..898
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 100
;   OTHER INFORMATION: region 270..325
;   OTHER INFORMATION: id H50235
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 769..812
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 93
;   OTHER INFORMATION: region 358..401
;   OTHER INFORMATION: id H50235
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 808..844
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 100
;   OTHER INFORMATION: region 325..361
;   OTHER INFORMATION: id H50235
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 143..309
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 100
;   OTHER INFORMATION: region 141..307
;   OTHER INFORMATION: id H06750
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 2..144
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 100
;   OTHER INFORMATION: region 1..143
;   OTHER INFORMATION: id H06750
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 306..375
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 91
;   OTHER INFORMATION: region 305..374
;   OTHER INFORMATION: id H06750
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 777..1057
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 100
;   OTHER INFORMATION: region 103..383
;   OTHER INFORMATION: id AA618048
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 1066..1161
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 98
;   OTHER INFORMATION: region 1..96
;   OTHER INFORMATION: id AA618048
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 454..636
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 97
;   OTHER INFORMATION: region 104..286
;   OTHER INFORMATION: id H20386
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 384..455
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 97
;   OTHER INFORMATION: region 35..106
;   OTHER INFORMATION: id H20386
```

```
;
; FEATURE:
;   NAME/KEY: est
;   LOCATION: 626..681
;   IDENTIFICATION METHOD: blastn
;   OTHER INFORMATION: identity 96
;   OTHER INFORMATION: region 277..332
;   OTHER INFORMATION: id H20386
;
; US-60-081-563-74
;
; Query Match
; Best Local Similarity 100.0%; Score 1173; DB 72; Length 1173;
; Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 gagctgttggagacaccgcttctcgcgccctcttaacgctgctgctgcgcgcgcca 60
;   |||||||
; Db 1 GAGCTGCTTATGGGACACCGCTTCTCGCGGGCCTCTTAAAGCTGCTGCTGCCGCCGCA 60
;   |||||||
; Qy 61 cccctgtatacccgccaccgcatgctcggtccagagtcgctcccgcccccaaaacgatac 120
;   |||||||
; Db 61 CCCCTGTATACCGGCACCGCATGCTCGGTCCAGATCCGTCGCCGCCCAACACGATCC 120
;   |||||||
; Qy 121 qcagcaaaactcatgaccccccgcgaatcgagcgacacaaatggacacattccactcgac 180
;   |||||||
; Db 121 CGCAGCAAACTCATYGGCACCGCCCGCAATCGGACGCAATGGCACCTTCCACTCGGAC 180
;   |||||||
; Qy 181 gaggcactggcatgcgcaactgcttcgctcctgcccggagtagccgggatgcagagattgtg 240
;   |||||||
; Db 181 GAGGCATGGCATGGCATGCTTCCGCTCCTGCGGGAGTAGCCGGGATGCGAGAGATTGTG 240
;   |||||||
; Qy 241 cggaccgggattcccgaaaaactcgcttcctgtgacatcggtgagcgtggtggggcgag 300
;   |||||||
; Db 241 CGGACCGGGATCCCGAAAACTCCTTCTGTGACATCGTGGTGAGACGTGGGGGGCGAG 300
;   |||||||
; Qy 301 tagaccctcggagacaccgatgacccatcacccagaggtctttcacagagaccatgagc 360
;   |||||||
; Db 301 TAGACCCCTCGGAGACACCGATATGACCATCACAGAGGTCTTTTCACAGAGACCATGAGC 360
;   |||||||
; Qy 361 tccctgtccctcggagggcggtgagaccgaactgtgagcagtgcggggactcatctctg 420
;   |||||||
; Db 361 TCCTGTCCCTGGGAGGCGGTGGCAGACCAAGCTGAGCAGTGGCGGACTCATCTATCTG 420
;   |||||||
; Qy 421 cacttcgggcacaagctgctgcccagtgctggcacttagtgaagagagacagcagtggtg 480
;   |||||||
; Db 421 CACTTCGGGCACAAGCTGCTGCCACAGTTGCTGGGCACCTAGTGAAGAGAGACAGCATGTG 480
;   |||||||
; Qy 481 ggcacctctatgacaagatgtatgagaactttgtggaggggtgagtgctgtggacaat 540
;   |||||||
; Db 481 GGCACCTCTATGACAAAGATGTATGAGAACTTTGTGGAGAGGTGGATGCTGTGACAAAT 540
;   |||||||
; Qy 541 gggatctccagtgggcagaggggagcctcgatgctgactgaccactaccctgagtgca 600
;   |||||||
; Db 541 GGGATCTCCAGTGGCAGAGGGGGAGCCCTCGATATGCACTGACCCACTACCTGAGTGCA 600
;   |||||||
; Qy 601 cgagtgtcgtacttaactcctacgtgaaaccccccgcgaacacacactgagcaggggttc 660
;   |||||||
; Db 601 CGAGTTGCTCGACTTAATCCTACTTGAACACACCCCGCAACAGACACTGAGGCGAGGTTC 660
;   |||||||
; Qy 661 aagcgtgcaatggatctggttcaagaggagttcttcagagattagattcttaccacac 720
;   |||||||
; Db 661 AAGCGTGCAATGGATCTGGTTCAAGAGGAGTTTCTGCAGAGATTAGATTTCATCCACAC 720
;   |||||||
; Qy 721 agctggctccagcccgccgttgggtggaagagcccttgcagcgattccaggtggag 780
;   |||||||
; Db 721 AGCTGGCTCCAGCCCGGCGCTTGGTGGAGAGAGGCCCTTGCCCCAGCGATTTCAGGTGGAC 780
;   |||||||
; Qy 781 ccaagtggagagattgtggaactggcgaagggtgcattgcctctggaaaggagcattctac 840
;   |||||||
; Db 781 CCAAGTGGAGAGATTGTGGAACCTGGCGAAAGTGCAATGTCCCTCGAAGAGGACATCTCTAC 840
;   |||||||
; Qy 841 cacttgaactctggctgtccctcccagtgccatctcttcttattctacactgaccag 900
;   |||||||
; Db 841 CACCTGGAATCTGGGCTGTCCCTCCAGTGGCCATCTCTCTTTGTTATCTACACTGACCAG 900
;   |||||||
```

```
QY 901 gctggacagtggcgaatacacagtgtgtcccaagagagcccaactattccaaagcgcggtg 960
Db 901 GCTGGACAGTGGCGAATACAGTGTGTGCCCAAGAGAGCCCACTCATTTCCAAAGCGGCGTG 960
QY 961 cccctgcagagccatggcggggtctcttcggagcagagccctgagccaggttcagtggatc 1020
Db 961 CCCCTGCAGAGCCATGGCGGGGTCTTCGGAGCAGAGCCCTGGACCAGGTCAGTGGGATC 1020
QY 1021 cctggctgcattcttcccatgaagcggttcattgagcgttcaccgcacccagagaggt 1080
Db 1021 CCTGGCTGCATCTTCGTCATGCAAGCGGCTTCATTGGCGGTTCACCCGACCCGAGAGGGT 1080
QY 1081 gctttgagcatggccgtgcacattggccagcgctcatacctccacaaaatctctag 1140
Db 1081 GCCTTGAGCATGGCCGTGCCACTTGGCCAGCGCTCATACCTCCACAAATCTCTCTAG 1140
QY 1141 tctaataaaaccttccatctcctcaaaaaa 1173
Db 1141 TCTAATAAAACCTTCCATCTCAAAAAA 1173

RESULT 3
US-09-515-126-14066/c
; Sequence 14066, Application US/09515126
; GENERAL INFORMATION:
; APPLICANT: Taq, Yuanhua T.
; APPLICANT: Tillinghast, John
; APPLICANT: Sinku, Ankura
; APPLICANT: Liu, Chenghua
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Contigs Obtained
; FILE REFERENCE: 788
; CURRENT APPLICATION NUMBER: US/09/515,126
; CURRENT FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 14074
; SOFTWARE: pl_CT_genes Version 1.2
; SEQ ID NO 14066
; LENGTH: 1456
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1456)
; OTHER INFORMATION: n = a,t,c or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1032)...(69)
; OTHER INFORMATION: similar to gi3874563 in the genepept database release 114,
; OTHER INFORMATION: Run with FASTX 3.3t00, default parameters
US-09-515-126-14066

Query Match 99.7%; Score 1169.8; DB 57; Length 1456;
Best Local Similarity 99.8%; Pred. No. 4.1e-284;
Matches 1171; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagctgttatggacaccgtctctcgcgccctttaacgtgtgctgctgcgcgcgcca 60
Db 1175 GAGCTGCTTATGGACACCGCTTCTCGGGGCTCTTAAGCTGTCTGTCTGCGCGGCCA 1116
QY 61 cccctgtataccggaccgcatgctcggtccagagtcggtcccgcccccaaacgatcc 120
Db 1115 CCCCTGTATACCGGACCGCATGCTCGGTCCAGAGTCGCTCCCGCCCCCAACGATCC 1056
QY 121 cgaagcaaacatcaggcaccgcccgaatcgggagcgacacaaatggcacttccactgcgac 180
Db 1055 CGCAGAAAACATCATGGACCGCCCGCGAATCGGGACGCAACAATGGCACTTCCACTGCGAC 996
QY 181 gagcactggcatcgcaactgctctcctcctgcgcgagtagtcaggagattgtg 240
Db 995 GAGGCACTGGCATCGCACGTGCTTCGCTCTCTGCGGAGTACCGGGATGCGAGATTGTG 936
```

```
QY 241 cggaccgggatcccgaataactcgcttctctgtacatcgtgggtgagcgtggggggcgag 300
Db 935 CGGACCCGGGATCCCGAAAAACTCGCTTCTGTGCATCGTGGTGGACGCTGGGGGGCGAG 876
QY 301 taagacctcggagacaccgatatgaccatcaccagaggtctcttcacagagaccatgagc 360
Db 875 TACGACCTCGGAGACACACCGATATGACCATCACCAGAGGTCTTTTCACAGAGACCATGAGC 816
QY 361 tccctgtccctcggagagcggtggcagacaaagctgagcagtgagcgggactatctatctg 420
Db 815 TCCCTGTGCCCTTGGGAAGCCGTGGCAGACCAAGCTGAGCAGTGGGGACTCATCTATCTG 756
QY 421 cactcgggcacaaagctgctggcccagttgctgggcactagtgaagagagacagcatggtg 480
Db 755 CACTTCGGGCACAACGCTGCTGGCCCACTGCTGGGCACCTAGTAGAGAGGACAGCATGGTG 696
QY 481 ggcacctctatgacaagatgtatgaaactttgtgaggggtgggtgctgtgtgacaat 540
Db 695 GGCACCTCTATGACAAAGATGTATGAGAACTTTGTGGAGGAGGTGGATGCTGTGGACAAT 636
QY 541 gggatctccagtgggcagagggggagcctcgatgacactgacacactaccctcagtgca 600
Db 635 GGGATCTCCAGTGGGAGAGGGGGAGCCTCGATATGCACTGACCTACCTTACCTGAGTGCA 576
QY 601 cgagtgtcgacttaactcctacctgaaacccaccacaaagacacactgagcaggggttc 660
Db 575 CGAGTTGCTGCACTTAATCTTACCTGGNACCACCCGCCAAGACACTGAGCGAGGGTTTC 516
QY 661 aagcgtgcaatggatcgtgggttcaagagaggtttctcgcagagattagatttctacaacac 720
Db 515 AAGCGTGCAATGGATCTGGTTCAAGAGGAGTTTCTGCAGAGATTAGATTCTTACCAACAC 456
QY 721 agctggctgcagcccgccgttgggtggaagagcccttggcccgccgacttccaggtggag 780
Db 455 AGCTGGCTGCCAGCCCGGCCCTTGGTGGAGAGGCCCTTGCCCGAGCATTCAGGTGGAC 396
QY 781 ccaagtggagagattgtggaactgacgaaggtgcattgctcctggaagagacatctctac 840
Db 395 CCAAGTGGAGAGATTGTGGAACCTGGCNAAGTGCAATGTCCCTGGAAGGAGCATCTCTAC 336
QY 841 cacctggaacttggtgctgctccctccagtgccatcttcttgttatctacactgaccag 900
Db 335 CACCTGGAATCTGGGCTGTCCCTCCAGTGGCCATCTCTTTGTATCTACACTGACCAG 276
QY 901 gctggacagtggcgaatacacagtgtgtgccccagagagccccactcattccaaagcggctg 960
Db 275 GCTGGACAGTGGCGAATACAGTGTGTGCCAAGGAGCCCCACTCATTCCAAAGCGCGCTG 216
QY 961 cccctgcagagccatggcggggtcttcggagcagagccctggaccaggtcagtgggatc 1020
Db 215 CCCCTGCCAGAGCCATGGCGGGGTCTTCGGGACAGAGGCCCTTGGACAGGTGAGTGGGATC 156
QY 1021 cctggctgcattctcctcatgcaagcggttcattggtcaccgcacccagagaggt 1080
Db 155 CCTGGCTGCATCTTCGTCCATGCAAGCGGCTTCATTGGCGGTCAACCGACCCGAGAGGGT 96
QY 1081 gccttgagcatggccgtgcccacttggcccagcgcctacaccccccaaatctcttag 1140
Db 95 GCCTTGAGCATGGCCGTGGCCACCTTGGCCCCAGCGCTCATACCTCCCAACAAATCTCTCTAG 36
QY 1141 tctaataaaaccttccatctcctcaaaaaa 1173
Db 35 TCTAATAAAACCTTCCATCTCAAAAAA 3
```

RESULT 4
PCT-US00-08981-23
; Sequence 23, Application PC/TUS00008981
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 48 Human Secreted Proteins
; FILE REFERENCE: P5547PCT

; CURRENT APPLICATION NUMBER: PCT/US00/08981
; CURRENT FILING DATE: 2000-04-06
; EARLIER APPLICATION NUMBER: 60/128,701
; EARLIER FILING DATE: 1999-04-09
; EARLIER APPLICATION NUMBER: 60/177,166
; EARLIER FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 154
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1225
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1214)
; OTHER INFORMATION: n equals a,t,g, or c
PCT-US00-08981-23

Query Match 99.7%; Score 1169.4; DB 1; Length 1225;
Best Local Similarity 99.7%; Pred. No. 4.9e-284;
Matches 1170; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gagctgttatggagacaccgttctctcgtcgccgtctttaaagctgtctgctgcgcgcgcca 60
Db 38 gagctgttatggagacaccgttctctcgtcgccgtctttaaagctgtctgctgcgcgcgcca 97
Qy 61 cccctgtatacccgccacccgcatgctcggtccagagtcctccgcccccaaacgacac 120
Db 98 cccctgtatacccgccacccgcatgctcggtccagagtcctccgcccccaaacgacac 157
Qy 121 cgcagcaaaactcatggaccccccgaatcgggacgcacaaatggccacttccactcgac 180
Db 158 cgcagcaaaactcatggaccccccgaatcgggacgcacaaatggccacttccactcgac 217
Qy 181 gagcaactggcatcgcaactgtctcgtcctcgcggagtagccgggagtagagattgtg 240
Db 218 gagggcactggcatcgcaactgtctcgtcctcgcggagtagccgggagtagagattgtg 277
Qy 241 cggaccggggtcccgcaaaaactcgtctcgtgacatcgtggtggcgcgtggggggcgag 300
Db 278 cggaccggggtcccgcaaaaactcgtctcgtgacatcgtggtggcgcgtggggggcgag 337
Qy 301 tacacactcgagacacccgatatgaccatcaccagaggtttttccacagagaccatgagc 360
Db 338 tacgaccctcgagacacccgatatgaccatcaccagaggtttttccacagagaccatgagc 397
Qy 361 tccctgtccctcgtggagggccgtggcagaccagctgagcagtgcgggactcatctatctg 420
Db 398 tccctgtccctcgtggagggccgtggcagaccagctgagcagtgcgggactcatctatctg 457
Qy 421 cacttcgggcaacaactgctggccagttgctggcactagtgaagaggacagcatggtg 480
Db 458 cacttcgggcaacaactgctggccagttgctggcactagtgaagaggacagcatggtg 517
Qy 481 ggcacctctatgacaagatgatgagaactttgtggaggaggtggatgctgtggacaat 540
Db 518 ggcacctctatgacaagatgatgagaactttgtggaggaggtggatgctgtggacaat 577
Qy 541 gggatctccagtgggaggggggagctcgatgacactgaccactaacctgagtgca 600
Db 578 gggatctccagtgggaggggggagctcgatgacactgaccactaacctgagtgca 637
Qy 601 cgagttgtcgtacttaactcctaccctggaaaccccccgaacacagacactgaggggttc 660
Db 638 cgagttgtcgtacttaactcctaccctggaaaccccccgaacacagacactgaggggttc 697
Qy 661 aagcgtgcaatggatcgtgttcaagaggagtttctgcagagagattagatttaccacac 720
Db 698 aagcgtgcaatggatcgtgttcaagaggagtttctgcagagagattagatttaccacac 757
Qy 721 agctggtccagcccgcccttggtggagggcccttgccagcgattccagtgagac 780

Db 758 agctggctgccagcccgggaccttggtggaagagggcccttgcccagcgattccaggtgac 817
Qy 781 ccaagtgagagagattgtggaactggcgaagagtgcatgtccctggaagagagcatctctac 840
Db 818 ccaagtgagagagattgtggaactggcgaagagtgcatgtccctggaagagagcatctctac 877
Qy 841 cactggaatctggcgtgctccctccagtgcccatctctttttgttatctacactgaccag 900
Db 878 cactggaatctggcgtgctccctccagtgcccatctctttttgttatctacactgaccag 937
Qy 901 gctggacagtggcgaatacatagtgtgtgccaaagagagcccaactcattccaaagcggctg 960
Db 938 gctggacagtggcgaatacatagtgtgtgccaaagagagcccaactcattccaaagcggctg 997
Qy 961 cccctgcagagccatggcgggtcttcggacgagccctggaccaggtcagtgggatc 1020
Db 998 cccctgcagagccatggcgggtcttcggacgagccctggaccaggtcagtgggatc 1057
Qy 1021 cctggctcatcttctcgtcgaagcggcttcatgtggtgcacccgaccccgagaggggt 1080
Db 1058 cctggctcatcttctcgtcgaagcggcttcatgtggtgcacccgaccccgagaggggt 1117
Qy 1081 gcttgagcatggcccggtgcacacttgcccaagcgtcctacactcccaaatctcctag 1140
Db 1118 gcttgagcatggcccggtgcacacttgcccaagcgtcctacactcccaaatctcctag 1177
Qy 1141 tctaataaaaccttccatctcaaaaaa 1173
Db 1178 tctaataaaaccttccatctcaaaaaa 1210

RESULT 5

US-09-474-436-8324
; Sequence 8324, Application US/09474436
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Fraser, Christopher C.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Donovan, Michael J.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1075001
; CURRENT APPLICATION NUMBER: US/09/474,436
; CURRENT FILING DATE: 1999-12-28
; EARLIER APPLICATION NUMBER: 60/113,970
; EARLIER FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 60/126,974
; EARLIER FILING DATE: 1999-03-30
; EARLIER APPLICATION NUMBER: 60/131,613
; EARLIER FILING DATE: 1999-04-29
; NUMBER OF SEQ ID NOS: 8332
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8324
; LENGTH: 1402
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-474-436-8324

Query Match 99.3%; Score 1165; DB 56; Length 1402;
Best Local Similarity 99.6%; Pred. No. 6.5e-283;
Matches 1168; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gagctgttatggagacaccgttctcgtcgggcctttaaagctgtctgctgcgcgcgcca 60
Db 160 gagctgttatggagacaccgttctcgtcgggcctttaaagctgtctgctgcgcgcgcca 219
Qy 61 cccctgtatacccgccacccgcatgctcgtgtccagagtcctccgcccccaaacgacac 120
Db 220 cccctgtatacccgccacccgcatgctcgtgtccagagtcctccgcccccaaacgacac 279
Qy 121 cgcagcaaaactcatggcaccgcccgccgaatcgggacgcacaaatggccacttccactcgac 180

```
Db 280 cgagcaaaactcatggcaccgcccggaatcgggacgcacaaatggcaacttccactgcgac 339
QY 181 gagcactggcatgcgcactgcttcctcctcctccggagtaaccggaatcgagagattgtg 240
Db 340 gaggcactggcatgcgcactgcttcctcctcctccggagtaaccggaatcgagagattgtg 399
QY 241 cgagcccggtatcccgaaaaactcgcttcctctgacatcgctggtggagctgggggagcgag 300
Db 400 cgagcccggtatcccgaaaaactcgcttcctctgacatcgctggtggagctgggggagcgag 459
QY 301 tagacacctcgagagacaccgatagaccatcacagagagttctttacagagacacatgagc 360
Db 460 tagacacctcgagagacaccgatagaccatcacagagagttctttacagagacacatgagc 519
QY 361 tcctgtccctcgagggcgctggagacacaaagctgagcagtgctgggactcatctatctg 420
Db 520 tcctgtccctcgagggcgctggagacacaaagctgagcagtgctgggactcatctatctg 579
QY 421 cacttggggcacaaactgctggcccgagttgctggcaactagtgaagaggacagcatggtg 480
Db 580 cacttggggcacaaactgctggcccgagttgctggcaactagtgaagaggacagcatggtg 639
QY 481 ggaacctctatgacaagatgtagagaaacttgtgagagaggtggtatgtgtggacaat 540
Db 640 ggaacctctatgacaagatgtagagaaacttgtgagagaggtggtatgtgtggacaat 699
QY 541 ggaacctcccgagggcgagagggagcctcgatatgacatgacacacacacacacacacac 600
Db 700 ggaacctcccgagggcgagagggagcctcgatatgacatgacacacacacacacacacac 759
QY 601 cgagttgctgacttaactctactgctggaacacaccccgacacacacacacacacacacac 660
Db 760 cgagttgctgacttaactctactgctggaacacaccccgacacacacacacacacacacac 819
QY 661 aagcgtgcaatgagatcgttgaagaggagttcttctgcagagattagattctaccacac 720
Db 820 aagcgtgcaatgagatcgttgaagaggagttcttctgcagagattagattctaccacac 879
QY 721 agctggctgcagccggcgcttggtagagagcccttgcacagccttccacagattccaggtgac 780
Db 880 agctggctgcagccggcgcttggtagagagcccttgcacagccttccacaggtggac 939
QY 781 ccaagtggagagatttggaaactggcgaaagggtgcatgtccctggaaaggagcatctctac 840
Db 940 ccaagtggagagatttggaaactggcgaaagggtgcatgtccctggaaaggagcatctctac 999
QY 841 cacttgaatctgggctgtccctccagtgccacatcttcttcttatctacactgaccag 900
Db 1000 cacttgaatctgggctgtccctccagtgccacatcttcttcttatctacactgaccag 1059
QY 901 gctggacagtgccgaatacacagtgtgtcccaaggagcccttccacacattccaaagccgctg 960
Db 1060 gctggacagtgccgaatacacagtgtgtcccaaggagcccttccacacattccaaagccgctg 1119
QY 961 cccctgcagagaccatgctgggggtcttcgggacgagagcccttggaccaggttcagtgggatc 1020
Db 1120 cccctgcagagaccatgctgggggtcttcgggacgagagcccttggaccaggttcagtgggatc 1179
QY 1021 cctggctgcattcttcgtccatgcaagcgggttcttcaattggcggttccacgcacccagagaggt 1080
Db 1180 cctggctgcattcttcgtccatgcaagcgggttcttcaattggcggttccacgcacccagagaggt 1239
QY 1081 gcttggagatggcccgctgcacacttggccagcgctcatacctcccaaaaatctctag 1140
Db 1240 gcttggagatggcccgctgcacacttggccagcgctcatacctcccaaaaatctctag 1299
QY 1141 tctaataaaaccttccatctcaaaaaa 1173
Db 1300 tctaataaaaccttccatctcaaaaaa 1332
```

RESULT 6

US-08-933-750-74

```
; Sequence 74, Application US/08933750
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LINDNOT03
; CLONE: 1572888
; US-08-933-750-74
```

```
Query Match 99.2%; Score 1163.4; DB 28; Length 1380;
Best Local Similarity 99.5%; Pred. No. 1.6e-282;
Matches 1167; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 gagctgtattggacaccgcttctcgtcgggcctcttaacgctgtcgtcgcgcgcga 60
Db 46 GAGCTGCTTATGGACACCGCGTCTCTGGCGGCCCTCTTAACGCTGCTGCTGCCCGCGCA 105
QY 61 cccctgtataccggcaccgcatgctcgggtccagagtcctcccgcccccaaacgatcc 120
Db 106 CCCCTGTATACCGGCACCGCATGCTCGGTCCAGAGTCCGTCGCCGCCCAAAACGATCC 165
QY 121 cgaagaaaactcatggcaccgccccgaatcgggacgcacaaatggcaaccttccactgcagc 180
Db 166 CGCAGCAAACTCATGTGGACCCGCCGAATCGGAGCGCACATATGCGACATGCCCTTCCACTGCGAC 225
QY 181 gagcactggcatgcgactgcttctcctcctgcgagtagccggatgcagagattgtg 240
Db 226 GAGCACTGGCATGCCACTGCTCTCGGAGTACCGGATACCGGATGCGAGATTGTG 285
QY 241 cggaccgggatacccgaaaaactcgttctctgtgacatcggtggagcgtgggggagag 300
Db 286 CGGACCGGGATCCCGAAAAACTCGCTTCCTGTGACATCGTGGTGGAGCTGGGGGGGAG 345
```

Wed May 24 09:19:44 2000

us-09-215-435-117.rnp

APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LINDNOT03
CLONE: 1572888
US-09-234-613-74

301 tacgacctcggagacacgatatgacacacagaggtttttcacagagacatgagc 360
346 TAGGACCCCTGGAGACACCGATATGACCATACACAGAGGTCTTTCACAGACCATGAGC 405
361 tcctgtcccttggaggccgtggcagaccagctgagcagtgagggtgacttatctatg 420
406 TCCCTGTCCCTGGGAAGCCGTGGCAGACCAAGCTGAGCAGTGCAGGACTCATCTATCTG 465
421 cacttggcagcaagctgtgcccagctgtggcagctgagcagctgagcagcagcagctg 480
466 CACTTGGGCACAAAGCTGTGGCCCAAGTTGCTGGGCACTAGTGAAGAGCAGCATGTG 525
481 ggcacctctatgacaaagatgatgagaactttgtggaggaggtgagctgtggaacaat 540
526 GGCACCCCTCTATCACAAAGATGATGAGAACTTTGTGGAGGAGGTGATGTGTGACAAAT 585
541 gggatctccagtgggcagagggggagccctcgatatgcaactgacacacacccctgagcga 600
586 GGGATCTCCCAAGTGGCAGAGGGGAGCCCTCGATATGCACTGACCACTACCTGAGTGA 645
601 cgaattgtctgacttaactctactctggaaccccccagccagacacactgagcaggttc 660
646 CGAGTTGCTCGACTTAATCTACTTGGACCCACCCGACCAAGACACTGAGCGAGGGTTC 705
661 aagcgtgcaatgagctggttcaagaggaggttcttcgacagagattagatttaccacaac 720
706 AAGCGTGCAATGGATCTGGTTCAAGAGGAGTTTCTGCAGAGATTAGATTCTACCAACAC 765
721 agctggctgcccagccggcctgtggaagagggccttgcacagcagattccagggtgac 780
766 AGCTGGCTGCCAGCCCGGGCTTGGTGGGAAGAGGCCCTTGCACAGCATTCACAGGTGAC 825
781 ccaagtggagagattgtggaactggcgaagaggtgcatgtccctgggaagagacatctctac 840
826 CCAAGTGGAGAGATTGTGGAATCGCCGAAGGTGCATGTCCTTGGAGGAGCATCTCTAC 885
841 cacctggaatctggctgtccctccagtgccatcttctgttatctacactgaccag 900
886 CACCTGGAATCTGGGCTGTGCCCCCTCCAGTGGCCATCTCTTTGTTATCTACATGACCAG 945
901 gctgacagtgccgaatacacagtgtgcccgaagagagcccaactcattcccaagccggctg 960
946 GCTGGACAGTGGCGAATACAGTGTGCCCCAAGGAGGCCCACTATTCCTCAAGCCGGCTG 1005
961 cccctggcagagcagtgccgggttcttgggagcagagccctggacacaggtcagtgagtc 1020
1006 CCCCCTGCCAGAGCCATGCGGGGCTTGGGAGCAGAGGCCCTTGGACCAAGGTGAGGATC 1065
1021 cctgtgcatcttctgcatgcaagcggcttcttggcagcagcctgagcagcagcagaggg 1080
1066 CTTGGCTGCATCTTCTGTCATGCAAGCGGCTTCATTGGCGGTCAACCCAGAGAGGGT 1125
1081 gcttggagcagtgccgtgcccaccttggccagcgcgtcattccacacaaatctctag 1140
1126 GCCTTGAGCATGGCCGCTGCCACCTTGGCCAGGCGCTCATACCTCCACAAATCTCTAG 1185
1141 tctaataaaccttccatctcaaaaaa 1173
1186 TCTAATAAAACCTTCCATCTACTACTGACAAA 1218

RESULT 7
US-09-234-613-74
; Sequence 74, Application US/09234613
; GENERAL INFORMATION:
; APPLICANT: Lai, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.

Query Match 99.2%; Score 1163.4; DB 44; Length 1380;
Best Local Similarity 99.5%; Pred. No. 1.6e-282;
Matches 1167; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gagctgtcttatgggacacgcgttctctgcgcggcctcttaacgctgctgctgccgcgcga 60
DB 46 GAGCTGCTTATGGGACACCGCTTCTTGGCGGGCTCTTAACGCTGCTGCTGCCGCCGCA 105
QY 61 cccctgtatacccgccagcagcagctgctggtccagagtcgctcccgcccccacaaacgaccc 120
DB 106 CCCCCTGTATACCGGGACCGCATGCTCGGTCCAGAGTCCGTCCCGCCGCCCAACGATCC 165
QY 121 cgcagcaaaactatgacaccccccgaatcgggagcagcaaatggcaccttccactgcgac 180
DB 166 CGCAGCAAACTCATGGCACCGCCCGAATCGGAGCACAATGGCACCTTCCACTCGCAG 225
QY 181 gaggcactggcagcagcagctgttcctcctgcgcggagctaccggagtcagagattgtg 240
DB 226 GAGGCATGCGCATGCGCACTGCTTCCCTCTCGCGGAGTACCGGATGCAGAGATTGTG 285
QY 241 cggacccggagatcccgaaaaactcgttctctgtgacatcgtgtgagcgtggggggcgag 300
DB 286 CGGACCCGGGATCCCGAAAACTCGCTTCTCTGTGACATCGTGTGGACGTGGGGGCGGAG 345
QY 301 tacgacctcggagacacgatatgacacagaggttcttcacagagacatgagc 360
DB 346 TAGACCCCTCGAGACACCGATATGACCATACCAAGAGGTCTTTCACAGAGACCATGAGC 405
QY 361 tcctgtcccttggaggccgtgagcagcaagctgagcagtcgagcagtcactctatctg 420
DB 406 TCCCTGTCCCTTGGGAAGCCGTGGCAGAGCCCAAGCTGAGCAGTGCAGGACTCATCTATCTG 465

QY 421 cacttggggcacaagctgtgccccagttgtctgggccaactagtgaagagagacagcatggtg 480
 Db 466 CACTTCGGGCGACAAAGCTGCTGGCGCCAGTTGCTGGGCACTAGTGAAGAGGACAGCATGGTG 525
 QY 481 ggcacctctatcacaaagtgtatgagaactttgtggaggagtgatgctgtggacaat 540
 Db 526 GGCACCCCTCTATGACAAGATGTATGAGAACTTTGTGGAGAGGTGGATGCTGTGGACAAT 585
 QY 541 gggatctcccgatgggagagggggagcctcgatgtgacctgaccactaccctgagtgca 600
 Db 586 GGGATCTCCAGTGGGCGAGAGGGAGCCTCGATATGCATGCACTACCCCTGAGTGCA 645
 QY 601 cgagttgtctgacttaactcctctggaacccaccccgacccaaagacactgagggaggttc 660
 Db 646 CGAGTTGTCTCGACTTAATCTAGCTGGAACACACCCCGACCAAGACACTGAGGCGAGGGTTC 705
 QY 661 aagcgtgcaatgatctgttcaagaggagtttctgagagatttctgagatttctccaacac 720
 Db 706 AAGCGTGCAATGATCTGTGTTCAAGAGAGGATTTCTGCAGAGATTAGATTTCTACCAACAC 765
 QY 721 agctggctgcagcccgggccttgggtggagagggcccttgcacagcgatccaggtggac 780
 Db 766 AGCTGGCTGCACGCCGGGCTTGGTGGAGAGAGCCCTTGCCCGACCGATTCAGGTGGAC 825
 QY 781 ccaagtgagagattgtggaaactggcgaaaaggtgcatgtccctgggagggagcatctctac 840
 Db 826 CCAAGTGGAGAGATTGTGGAACTGGCGAAAGGTGCATGTCCCTGGGAAGGAGCATCTCTAC 885
 QY 841 cacttggaaactgggtgtccctcctcaagtgagcattcttcttattctacactgacag 900
 Db 886 CACTTGGAAATCTGGGTGTCTCCCTCCAGTGGCCATCTCTTTGTATCTACTAGACCAG 945
 QY 901 gctggacagtggcgaatcacagtgtgcccagagagcccaactcattccaaaagccgctg 960
 Db 946 GCTGGACAGTGGCGAATACAGTGTGTGCCACAGAGCCCCACATCTCCAAAGCCGCTG 1005
 QY 961 cccctgcagagccatggcggttcttctgggagagggccctggaccaggtcagtggtgac 1020
 Db 1006 CCCTGCGACAGGCATGGCGGGGTCTTCGGGACAGGSCCTGGACCAGGTGAGTGGGATC 1065
 QY 1021 cctggctgcatctctgtccatgcaagcggttctcattggcggtcacccgacccgagaggt 1080
 Db 1066 CTTGGCTGCATCTTCGTCCATGCAAGCGGTTTCATTGGCGGTTCACCGACCCGAGAGGGT 1125
 QY 1081 gcttgaagctgcccgtgccacttggcccagcgctcatacctccacaaatctcctag 1140
 Db 1126 GCCTTGAGCATGGCCCGTGCCACTTGCCCGAGGCTTCATACCTCCCAAAATCTCCTAG 1185
 QY 1141 tctaataaaaccttccatctcctcaaaaaa 1173
 Db 1186 TCTAATAAAACCTTCCATCTCATAGTACAAAA 1218

RESULT 8
 US-60-172-360-23018
 ; Sequence 23018, Application US/60172360
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, MacDonald
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Diep, Dinh
 ; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
 ; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor
 ; FILE REFERENCE: GX-0007 P
 ; CURRENT APPLICATION NUMBER: US/60/172, 360
 ; CURRENT FILING DATE: 1999-12-16
 ; NUMBER OF SEQ ID NOS: 29838
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 23018
 ; LENGTH: 1382
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:

Qy	960	gcccctgccagagccatggcggggtcttcgggacgagccctggaccaggtcaagtgggat	1019
Db	1007	gcccctgccagagccatggcggggtcttcgggacgagccctggaccaggtcaagtgggat	1066
Qy	1020	ccctggctgcattctgtccatgcaagcggtctcatgtgcggtcacgcgaccgcgagaggg	1079
Db	1067	ccctggctgcattctgtccatgcaagcggtctcatgtgcggtcacgcgaccgcgagaggg	1126
Qy	1080	tgccttgagcatggccgtggccaccttggccagcggtcatatctcccccaaatctctcta	1139
Db	1127	tgccttgagcatggccgtggccaccttggccagcggtcatatctcccccaaatctctcta	1186
Qy	1140	gtctataaaaccttcacatctcaaaaaaaaaaaaaa	1173
Db	1187	gtctataaaaccttcacatctcatactgacaaaa	1220

RESULT 9

US-09-359-922-3400

; Sequence 3400, Application US/09359922

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL CONFIGS OBTAINED FROM VARIOUS cdNA

; TITLE OF INVENTION: LIBRARIES

; FILE REFERENCE: 20411-752CON1

; CURRENT APPLICATION NUMBER: US/09/359,922

; CURRENT FILING DATE: 1999-07-22

; EARLIER APPLICATION NUMBER: US 09/205,155

; EARLIER FILING DATE: 1998-12-03

; NUMBER OF SEQ ID NOS: 13203

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 3400

; LENGTH: 1297

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-359-922-3400

```

RESULT 10
US-09-359-922-3400
; Sequence 3400, Application US/09359922A
;
; GENERAL INFORMATION:
;
; APPLICANT: Leshkowitz, Dena
;
; APPLICANT: Lib, Jin
;
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
;
; TITLE OF INVENTION: LIBRARIES
;
; FILE REFERENCE: 20411-752CON1
;
; CURRENT APPLICATION NUMBER: US/09/359,922A
;
; CURRENT FILING DATE: 1999-07-22
;
; EARLIER APPLICATION NUMBER: US 09/205,155
;
; EARLIER FILING DATE: 1998-12-03
;
; EARLIER APPLICATION NUMBER: US 09/034,341
;
; EARLIER FILING DATE: 1998-02-13
;
; NUMBER OF SEQ ID NOS: 13203
;
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 3400
;
; LENGTH: 1297
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
US-09-359-922-3400

```



```
Query Match          96.0%; Score 1125.8; DB 51; Length 1297;
Best Local Similarity 99.5%; Pred. No. 4.6e-273;
Matches 1171; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Qy 1 gagctgcttattggacacccgcttctcgtcgccgctctttaaagctgctgctgcgcgcga 60
Db 22 gagctgcttattggacacccgcttctcgtcgccgctctttaaagctgctgctgcgcgcga 81

Qy 61 cccctgtatacccgacccgcatgctcggttcagagtcgctcccgcccccaaaacgatcc 120
Db 82 cccctgtatacccgacccgcatgctcggttcagagtcgctcccgcccccaaaacgatcc 141

Qy 121 cgcagcaaaactcatggaccccccgaatcgggacgcacaaatggcaccttccactgcgac 180
Db 142 cgcagcaaaactcatggaccccccgaatcgggacgcacaaatggcaccttccactgcgac 201

Qy 181 gagcactggcatgcgcactgcttcgctcctcgtcgcggagtagccgggatgcagagattgtg 240
Db 202 gagcactggcatgcgcactgcttcgctcctcgtcgcggagtagccgggatgcagagattgtg 261

Qy 241 cggaccgggatcccgaaaaactcgtctcctgtgacatcgttggtggaagtggggcgag 300
Db 262 cggaccgggatcccgaaaaactcgtctcctgtgacatcgttggtggaagtggggcgag 321

Qy 301 tacgacctcgagacacccgatataccatcaccagagtgcttttcacagagaccatgagc 360
Db 322 tacgacctcgagacacccgatataccatcaccagagtgcttttcacagagaccatgagc 381

Qy 361 tccctgtccctggagccgctggcagaccagctgagcagtcgaggactcatctatctg 420
Db 382 tccctgtccctggagccgctggcagaccagctgagcagtcgaggactcatctatctg 441

Qy 421 cacttc-gggcacaaagctgctggccccagttcgtgggacctagtgaagagcagcatggt 479
Db 442 cacttcgggcacaaagctgctggccccagttcgtgggacctagtgaagagcagcatggt 501

Qy 480 -gggcacaccttatgacaagatgatgagaactttgtgagagagtgatgctgtggaca 538
Db 502 ggggcacaccttatgacaagatgatgagaactttgtgagagagtgatgctgtggaca 561

Qy 539 atgggatctccagtgggcagagggggagcctcgatatgactgacctaccctcagtg 598
Db 562 atgggatctccagtgggcagagggggagcctcgatatgactgacctaccctcagtg 621

Qy 599 cagagattgctcgaacttaactcactggaaccccccgcagacaaagacactgaggcaggt 658
Db 622 cagagattgctcgaacttaactcactggaaccccccgcagacaaagacactgaggcaggt 681

Qy 659 tcaagcgtgcaatggatctggttcaagaggaggtttcttcagagattagatttctaccaac 718
Db 682 tcaagcgtgcaatggatctggttcaagaggaggtttcttcagagattagatttctaccaac 741

Qy 719 acagctggctgccagcccgccctgttggaagagggcccttgcgcagcgattccagggtg 778
Db 742 acagctggctgccagcccgccctgttggaagagggcccttgcgcagcgattccagggtg 801

Qy 779 acccaagtggagagatttggaactggcgaaggttgcatgtcccttggaagagagatctct 838
Db 802 acccaagtggagagatttggaactggcgaaggttgcatgtcccttggaagagagatctct 861

Qy 839 accacctggaacttggtctccctccagtgggccatcttcttgttatctacactgacc 898
Db 862 accacctggaacttggtctccctccagtgggccatcttcttgttatctacactgacc 921

Qy 899 aggcctggacagtggcgaatacagtggtgcccgaagagcccccaactcattccaaagccggc 958
Db 922 aggcctggacagtggcgaatacagtggtgcccgaagagcccccaactcattccaaagccggc 981

Qy 959 tgcacctgccagagccatggcggggctcttcgggacgagggcccttgaccaggttcagtgga 1018
Db 982 tgcacctgccagagccatggcggggctcttcgggacgagggcccttgaccaggttcagtgga 1041
```

```
Qy 1019 tccctggctgcatcttgctccatgcaagcggtcttattggcggtcacccg-acccgagag 1077
Db 1042 tccctggctgcatcttgctccatgcaagcggtcttattggcggtcaacgcgacccgagag 1101

Qy 1078 ggtgcttgagcatggccgtgcccacttggcccagc-gctcatacctcccacaaatctc 1136
Db 1102 ggtgcttgagcatggccgtgcccacttggcccagcggtcatacctcccacaaatctc 1161

Qy 1137 ctactctaataaaaccttccatctcaaaaaaa 1173
Db 1162 ctactctaataaaaccttccatctcaaaaaaa 1198
```

RESULT 11

```
US-205-070-7647
; Sequence 7647, Application US/09205070
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-748
; CURRENT APPLICATION NUMBER: US/09/205,070
; CURRENT FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7647
; LENGTH: 1199
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-205-070-7647
```

Query Match 95.9%; Score 1124.8; DB 43; Length 1199;
Best Local Similarity 99.5%; Pred. No. 8.1e-273;
Matches 1170; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```
Qy 1 gagctgcttattggacacccgcttctcgtcgccgctctttaaagctgctgctgcgcgcga 60
Db 19 gagctgcttattggacacccgcttctcgtcgccgctctttaaagctgctgctgcgcgcga 78

Qy 61 cccctgtatacccgacccgcatgctcggttccagagtcgctcccgcccccaaaacgatcc 120
Db 79 cccctgtatacccgacccgcatgctcggttccagagtcgctcccgcccccaaaacgatcc 138

Qy 121 cgcagcaaaactcatggaccccccgaatcgggacgcacaaatggcaccttccactgcgac 180
Db 139 cgcagcaaaactcatggacccgcg-cccgaatcgggacgcacaaatggcaccttccactgcgac 197

Qy 181 gagcactggcatgcgcactgcttcgctcctcgtgacatcgttggtggaagtggggcgag 240
Db 198 gagcactggcatgcgcactgcttcgctcctcgtgacatcgttggtggaagtggggcgag 257

Qy 241 cggaccgggatcccgaaaaactcgtctcctgtgacatcgttggtggaagtggggcgag 300
Db 258 cggaccgggatcccgaaaaactcgtctcctgtgacatcgttggtggaagtggggcgag 317

Qy 301 tacgacctcgagacacccgatataccatcaccagaggtctttcacagagaccatgagc 360
Db 318 tacgacctcgagacacccgatataccatcaccagaggtctttcacagagaccatgagc 377

Qy 361 tccctgtccctggagccgctggcagaccagctgagcagtcgaggactcatctatctg 420
Db 378 tccctgtccctggagccgctggcagaccagctgagcagtcgaggactcatctatctg 437

Qy 421 cacttc-gggcacaaagctgctggcccagttgctgggacctagtgaagagcagcatggt 479
Db 438 cacttcgggcacaaagctgctggcccagttgctgggacctagtgaagagcagcatggt 497

Qy 480 -gggcacaccttatgacaagatgatgagaactttgtgagagagtgatgctgtggaca 538
Db 498 ggggcacaccttatgacaagatgatgagaactttgtgagagagtgatgctgtggaca 557

Qy 539 atgggatctccagtgggcagaggggagcctcgatatgactgacctaccctcagtg 598
```

Db	558	atgggtctccctcagtgaggaggggagccctcgatatactgactgaccactaccctcagtg	617
Qy	599	cacgagtgctgacttaatacctacctggaacacacccacacacactgagggaggggt	658
Db	618	cacgagtgctgacttaatacctacctggaacacacccacacacactgagggaggggt	677
Qy	659	tcaagcgtgcaatggatctggttcaagagaggtttctgcagagattagatttctaccaac	718
Db	678	tcaagcgtgcaatggatctggttcaagagaggtttctgcagagattagatttctaccaac	737
Qy	719	acagctgagctgcagcccgccctggttggaagagcccttgcacgagcattccaggtgg	778
Db	738	acagctgagctgcagcccgccctggttggaagagcccttgcacgagcattccaggtgg	797
Qy	779	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	838
Db	798	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	857
Qy	839	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	898
Db	858	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	917
Qy	899	aggtcgtgagcagtgccgagatacagtggtgcccagagagcccaactcattccaaagccgc	958
Db	918	aggtcgtgagcagtgccgagatacagtggtgcccagagagcccaactcattccaaagccgc	977
Qy	959	tgcccttgcagagagccatggcgggggttctgcgagagagcccttgcacaggtcagtgga	1018
Db	978	tgcccttgcagagagccatggcgggggttctgcgagagagcccttgcacaggtcagtgga	1037
Qy	1019	tccttggtgcattcttcctcatgcaagcggttcttctggtgcagcgc-acccgagag	1077
Db	1038	tccttggtgcattcttcctcatgcaagcggttcttctggtgcagcgcacccgagag	1097
Qy	1078	ggtgcctgagcagtgcccggtgcacacttggccagcgctcattacctcccaaaattcc	1137
Db	1098	ggtgcctgagcagtgcccggtgcacacttggccagcgctcattacctcccaaaattcc	1157
Qy	1138	tagtctaataaaccttccattctcaaaaaaaaaaaaa	1173

Db	558	atgggtctccctcagtgaggaggggagccctcgatatactgactgaccactaccctcagtg	617
Qy	599	cacgagtgctgacttaatacctacctggaacacacccacacacactgagggaggggt	658
Db	618	cacgagtgctgacttaatacctacctggaacacacccacacacactgagggaggggt	677
Qy	659	tcaagcgtgcaatggatctggttcaagagaggtttctgcagagattagatttctaccaac	718
Db	678	tcaagcgtgcaatggatctggttcaagagaggtttctgcagagattagatttctaccaac	737
Qy	719	acagctgagctgcagcccgccctggttggaagagcccttgcacgagcattccaggtgg	778
Db	738	acagctgagctgcagcccgccctggttggaagagcccttgcacgagcattccaggtgg	797
Qy	779	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	838
Db	798	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	857
Qy	839	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	898
Db	858	accacagtgagagagatttggaactggcgaaggtgcattgccttgcgaagagcattct	917
Qy	899	aggtcgtgagcagtgccgagatacagtggtgcccagagagcccaactcattccaaagccgc	958
Db	918	aggtcgtgagcagtgccgagatacagtggtgcccagagagcccaactcattccaaagccgc	977
Qy	959	tgcccttgcagagagccatggcgggggttctgcgagagagcccttgcacaggtcagtgga	1018
Db	978	tgcccttgcagagagccatggcgggggttctgcgagagagcccttgcacaggtcagtgga	1037
Qy	1019	tccttggtgcattcttcctcatgcaagcggttcttctggtgcagcgc-acccgagag	1077
Db	1038	tccttggtgcattcttcctcatgcaagcggttcttctggtgcagcgcacccgagag	1097
Qy	1078	ggtgcctgagcagtgcccggtgcacacttggccagcgctcattacctcccaaaattcc	1137
Db	1098	ggtgcctgagcagtgcccggtgcacacttggccagcgctcattacctcccaaaattcc	1157
Qy	1138	tagtctaataaaccttccattctcaaaaaaaaaaaaa	1173

RESULT 12

US-09-340-623-7647

: Sequence 7647, Application US/09340623

: GENERAL INFORMATION:

: APPLICANT: Hyseq, Inc.

: TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS cDNA

: TITLE OF INVENTION: LIBRARIES

: FILE REFERENCE: 20411-748CON1

: CURRENT APPLICATION NUMBER: US/09/340,623

: CURRENT FILING DATE: 1999-06-28

: EARLIER APPLICATION NUMBER: US 09/205,070

: EARLIER FILING DATE: 1998-12-03

: NUMBER OF SEQ ID NOS: 45207

: SOFTWARE: FastSEQ for Windows Version 3.0

: SEQ ID NO 7647

: LENGTH: 1199

: TYPE: DNA

: ORGANISM: Homo sapiens

US-09-340-623-7647

Query Match 95.9%; Score 1124.8; DB 50; Length 1199;

Best Local Similarity 99.5%; Pred. No. 8.1e-273;

Matches 1170; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Qy 1 gacgtcctatggagacacgcttctcgtcgccggtcttaacgctgtcgtgcgcgcgca 60

Db 19 gacgtcctatggagacacgcttctcgtcgccggtcttaacgctgtcgtgcgcgcgca 78

Query Match	57.8%	Score	677.6	DB	57	Length	1280
Best Local Similarity	89.6%	Pred. No.	2.6e-160	Indels	52	Gaps	31
Matches	1084	Conservative	0	Mismatches	74		

QY	1	gagctgcttatgggacacgcgttctctgcggcc- tcttaacgtctgtctgcgcgcgc	59
DB	54	gagctgcttatgggacacgcgttctctgcggccttctaacgtctgtctgcgcgcgc	113
QY	60	acccctgtatacccgccacccgcgtctgcgtctcagagatccgtcccgccccaacacgac	119
DB	114	acccctgtatacccgccacccgcgtctgcgtctcagagatccgtcccgccccaacacgac	173
QY	120	cgcagcaaaactatgacgcgcgcgcgaatcgggacgcacaaatggcaccttccactgcga	179
DB	174	cgcagcaaaactatgacgcgcgcgcgaatcgggacgcacaaatggcaccttccactgcga	233
QY	180	cgaggcaactggcatgcgcactcttcgcctcctgcggagtacgggatacagagattct	239
DB	234	cgaggcaactggcatgcgcactcttcgcctcctgcggagtacgggatacagagattct	293
QY	240	gcggaccgcggatcccgaaaaactc-gttcctctgacatcgtg- --gtgacgtggggg	295
DB	294	gcggaccgcggatcccgaaaaactcgtttcctctgacatcgtgggtgggagtggggggg	353
QY	296	gcgagtagcacc- tcggagacacgcgatatgacca- tcaccagaggtcttccacagac	353
DB	354	gcgagtagcaccctcggagacacgcgatatgaccattccacagaggtcttccacagac	413
QY	354	catgagctccctgctccctcggagggccgtgggcagaccacgtgcagcagtg- gggactca	412
DB	414	catgagctccctgctccctcgggaagccgtgggcagaccacgtgcagcagtgcggggactca	473
QY	413	tctatctgcacttc- ggggcacaagct- gctggccacgtgtgtggccactagtgaaagga	470
DB	474	tctatctgcacttcggggcacagctgggtggccacgtgtgtggccactagtgaaagga	533
QY	471	cagcatggt- gggcacctctatgacaagatgatgaaactttgtggaggaggtggatg	529
DB	534	cagcatggtggggcacccctctatgacaagatgatgaaactttgtggaggaggtggatg	593
QY	530	ctgt- ggacaatgggattctccagtgggc- agagggggagcctcgatatgcac- tgcacca	586
DB	594	ctgtgggacaatgggattctccagtgggcaagaggggagcctcgatatgcacttgacca	653
QY	587	ctaccctga- --gtgcacgagttgctgcac- ttaatcctaccctgggaacc- --accctga	638

	Query Match	43.3%	Score 507.8	DB 56	Length 572
	Best Local Similarity	96.8%	Pred. No. 1.1e-17		
	Matches 518	Conservative	Mismatches 17	Indels	Gaps
y	136	gcacgcgccgaatcgggacgcacacatggcactctccactgcgcgagcgactggcatgc	195		
b	38	gcttggcacgaggtggcgacgcacacatggcctctccactgcgcgagcgactggcatgc	97		
y	196	gcactgtctgcctcctgcgcgagtcacgcggatgcagagattgtcgcgacccgggatccc	255		
y	98	gcactctctgcctctgcgcgagtcacgcggatgcagagattgtcgcgagccgggatccc	157		

136	gcacgcgccgaatcgggacgacgacacaaatggcaccttccactgcacgagggcactggcagtc	195
QY		
38	gttggcacgaggtcggcgcgacaaatggcgcttccactgcacgagggcactggcagtc	97
Db		
196	gcactgttcgcctcctgcgcgagtagtcgggatgcagagattgtgcggaccggggatccc	255
QY		
98	ncactgttgcaccttcgcggatcaccgagatcagaagttgcgcgagggccgggattccc	157
Db		

Qy	256	gaaaaactcgtcttctgtgacatcgtgtgacgtggtggggtgaggtacgacctcggaga	315
Db	158	gaaaaactcgtcttctgtgacatcgtgtgacgtggtggggtgaggtacgacctcggaga	217
Qy	316	caccgatatgacctaccacagaggtctttcacagagacctagactccctgtccccctggg	375
Db	218	caccgatatgacctaccacagaggtctttcacagagacctagactccctgtccccctggg	277
Qy	376	aggcgtggcagacaaagctgagcagtcggtgacctctctatctgcacttcgggcacaag	435
Db	278	aagccgtggcagacaaagctgagcagtcggtgacctctctatctgcacttcgggcacaag	337
Qy	436	ctgctggcccagttgctgggcactagtgaaagagacagcatggtgggcacctctctatgac	495
Db	338	ctgctggcccagttgctgggcactagtgaaagagacagcatggtgggcacctctctatgac	397
Qy	496	aagatgtatgagaaactttgtggaggaggtggatgctgtggacaatgggatctccactgg	555
Db	398	aagatgtatgagaaactttgtggaggaggtggatgctgtggacaatgggatctccactgg	457
Qy	556	gcagagggggagcctcgatatgcactgacctaccctgagtgacagagttgctcgactt	615
Db	458	gcagagggggagcctcgatatgcactgacctaccctgagtgacagagttgctcgactt	517
Qy	616	aatcctacctggaaaccccccgcacacacactgagcgagggttcaagcgtgcaa	670
Db	518	aatcctacctggaaaccccccgcacacacactgagcgagggttcaagcgtgcaa	572

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 22:03:38 ; Search time 3267.04 Seconds
(without alignments)
1455.272 Million cell updates/sec

Title: US-09-215-435-117
Perfect score: 1173
Sequence: 1 gagctgcttatggacaccg.....tccatctcaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *

45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *
54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *
59: gb_est33: *
60: gb_est34: *
61: gb_est35: *
62: gb_est36: *
63: gb_est37: *
64: gb_est38: *
65: em_est27: *
66: em_est28: *
67: em_est29: *
68: em_est30: *
69: gb_est39: *
70: gb_est40: *
71: gb_est41: *
72: gb_est42: *
73: gb_est43: *
74: gb_est44: *
75: em_est31: *
76: em_est32: *
77: em_est33: *
78: em_est34: *
79: gb_est45: *
80: gb_est46: *
81: gb_est47: *
82: gb_gss1: *
83: gb_gss2: *
84: gb_gss3: *
85: gb_gss4: *
86: em_gss1: *
87: em_gss2: *
88: em_gss3: *
89: em_gss4: *
90: gb_gss5: *
91: gb_gss6: *
92: gb_gss7: *
93: gb_gss8: *
94: gb_gss9: *
95: em_gss5: *
96: em_gss6: *
97: em_gss7: *
98: em_gss8: *
99: em_gss9: *
100: em_gss10: *
101: em_gss11: *
102: gb_gss10: *
103: gb_gss11: *
104: em_gss12: *
105: gb_gss12: *
106: gb_gss13: *
107: gb_gss14: *
108: gb_gss15: *
109: gb_gss16: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 889 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 440.
Location/Qualifiers
1. .682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2366859"
/lab_host="DH108"
/lab_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/notes="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site: 1; Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneIDs: Soares NBHP pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

FEATURES
source

No.	Score	Match	Length	DB	ID	Description
C 1	673	57.4	682	62	AI911227	AI911227 wg33b02.x
C 2	624.8	53.3	695	70	AW149618	AW149618 xf40c06.x
C 3	605.2	51.6	610	51	AI760168	AI760168 wg58f03.x
C 4	602.6	51.4	609	64	AW025001	AW025001 wu70h04.x
C 5	583.6	49.8	635	64	AW082806	AW082806 xg03b01.x
C 6	539.4	46.0	541	50	AI700463	AI700463 wd14c04.x
C 7	536	45.7	537	79	AW303800	AW303800 xr21h10.x
C 8	529.6	45.1	555	51	AI752742	AI752742 cn18e03.x
C 9	526	44.8	536	62	AI936652	AI936652 wp68b08.x
C 10	506.4	43.2	558	62	AI928722	AI928722 w50e12.x
C 11	486.2	41.4	494	42	AI095901	AI095901 qb21b03.x
C 12	475.4	40.5	477	48	AI554935	AI554935 te48e12.x
C 13	473.2	40.3	492	42	AI138745	AI138745 qd98c10.x
C 14	442.8	37.7	600	45	AA818018	AA818018 ui-R-A0-a
C 15	430.8	36.7	435	44	AI277497	AI277497 ql96d08.x
C 16	430.2	36.7	445	40	AA972697	AA972697 op91d03.s
C 17	427.2	36.4	433	43	AI199598	AI199598 q159e05.x
C 18	425.8	36.3	431	43	AI241950	AI241950 qu68h06.x
C 19	420.4	35.8	485	45	AI357218	AI357218 qx17g04.x
C 20	419	35.7	424	81	AW452978	AW452978 ui-H-BW1-
C 21	417.2	35.6	446	61	AA830389	AA830389 wh9a07.x
C 22	412.6	35.2	512	24	H64785	H64785 yr58g05.r1
C 23	410.2	35.0	472	50	AI693289	AI693289 wd79a01.x
C 24	409.2	34.9	435	40	AA936697	AA936697 ol79e01.s
C 25	409.2	34.9	435	41	AI024270	AI024270 ov73h06.s
C 26	407.4	34.7	417	33	AA401345	AA401345 zu62d03.s
C 27	406.8	34.7	425	40	AA931135	AA931135 oo70b04.s
C 28	405.2	34.5	414	39	AA873221	AA873221 oh70g04.s
C 29	401.6	34.2	428	43	AI206725	AI206725 qf61c07.x
C 30	399	34.0	435	40	AA988748	AA988748 or75b03.s
C 31	398.6	34.0	405	40	AA988399	AA988399 og89h03.s
C 32	397.8	33.9	552	40	AA339993	AA339993 hc91c04.x
C 33	395.8	33.7	416	28	AA084391	AA084391 zn05f11.s
C 34	391.2	33.4	409	39	AA838228	AA838228 oe37d05.s
C 35	390.4	33.3	421	33	AA401585	AA401585 zu62d03.r
C 36	385.4	32.9	393	51	AI752743	AI752743 cn18e03.y
C 37	381.4	32.5	546	27	AA033451	AA033451 m142e07.r
C 38	378.8	31.8	475	27	AA033451	AA033451 m142e07.r
C 39	365.8	31.2	515	44	AI325965	AI325965 mq64e02.x
C 40	361.4	30.8	383	36	AA618048	AA618048 nq0f104.s
C 41	357.6	30.5	393	45	AI362439	AI362439 qv92h10.x
C 42	356	30.3	444	27	AA051437	AA051437 m141g10.r
C 43	356	30.3	475	23	H50235	H50235 yo28d03.s1
C 44	355.8	30.3	459	37	AA692410	AA692410 vt21d07.r
C 45	355.4	30.3	459	22	H06750	H06750 y183a04.r1

ALIGNMENTS

RESULT 1
AI911227/c mRNA EST 19-DEC-1999
LOCUS wg33b02.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2366859 3' similar to WP:K08H10.8 CE1880 YEAST HYPOTHETICAL
PROTEIN YE16 LIKE ;, mRNA sequence.
AI911227 682 bp
AI911227.1 GI:5630963
human.
ACCESSION Homo sapiens
VERSION Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
KEYWORDS Eutheria; Primates; Catarrhini; Hominidae; Homo.
SOURCE 1 (bases 1 to 682)
ORGANISM NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Unpublished (1997)
JOURNAL Tumor Gene Index
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948117.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

QY 1030 atcttcgctccatgcaagcggttcattggcggtcaccgacccgagaggggtgcttgagc 1089
 Db 142 ATCTTCCTCATGCAAGCGGCTTCATTGGCGGTTCACCGCACCCGAGAGGCTTGAGC 83
 QY 1090 atggccgctgccaaccttggcccgcgctctatactccacaaatctctctagtctaataa 1149
 Db 82 ATGGCCGCTGCCACCTTGGCCGCGCTCATAGCTCCACAAATCTCTAGTCTAATAAA 23
 QY 1150 accttcctcatcaaaaaaaa 1171
 Db 22 ACCTTCCATCTCATAAAAAAA 1

RESULT 2
 AW149618/c
 LOCUS
 DEFINITION
 xf40c06.x1 NCI_CGAP_Brn50 Homo sapiens cDNA clone IMAGE:2620522 3'
 similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL PROTEIN Y6 LIKE
 ;, mRNA sequence.

ACCESSION
 AW149618
 VERSION
 AW149618.1 GI:6197514
 KEYWORDS
 EST.
 SOURCE
 human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 695)
 NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 (CGAP/BTGP), Tumor Gene Index
 Unpublished (1998)
 On Jun 22, 1998 this sequence version replaced gi:3246919.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

COMMENT
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Burt Feuerstein, M.D., Mark Israel, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 465.
 Location/Qualifiers
 1..695
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2620522"
 /clone_lib="NCI_CGAP_Brn50"
 /tissue_type="medulloblastoma"
 /lab_host="DH10B (phage resistant)"
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 medulloblastoma tumor tissue, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. This library is normalized. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 151 a 186 c 200 g 157 t 1 others

BASE COUNT 151 a 186 c 200 g 157 t 1 others
 ORIGIN

Query Match 53.3%; Score 624.8; DB 70; Length 695;
 Best Local Similarity 96.4%; Pred. No. 5.8e-154;
 Matches 660; Conservative 0; Mismatches 23; Indels 2; Gaps 2;

QY 481 ggcacctctatgacaagatgagaaccttggaggaggtggatgctgtgacaat 540
 Db 691 GGCACCTATATACCAAGATGATAAGAAC-TTGTGTACAGCTGGATGCTGTGCAAT 633

QY 541 gggatctcccagtgggcagaggggagccctcgatatcactgaccactacccctgaagca 600
 Db 632 GGATTCFCCAGGGGGCAGAGGGCTTCGATATGCATGACCATCATTACCATGAGTGCA 573
 QY 601 c9agttgctcgacttaacctacctgaaacacacccacacacacacacacacacacacac 660
 Db 572 CGAGTTGCTCGACTTAATCATACCTGGAAACACCCCGACAAAGACACTGAGGCGAGGTC 513
 QY 661 aagcgtgcaatggtggttcaagagaggtttctgcagagatttagatttttaccacac 720
 Db 512 AAGCGTGCAATGGATCTGGTTCAGAGGAGTTCCTGAGAGATTAGATTCTTACAAACAC 453
 QY 721 agctggctccagcccggtcttggtggaagagcccttcccagcgatctcaggtgac 780
 Db 452 AGCTGGCTGCACGCCCGGGCTTGGTGAAGAGGCCCTTGGCCAGCAGATTCACAGTGGAC 393
 QY 781 ccaagtggagagatttggaactggcgaaggtgcatgtccctggaagagagcatctctac 840
 Db 392 CCAAGTGGAGAGATTGTGGAACCTGGCGAAAGGTGCATGTCCCTGGAAGGAGCATCTCTAC 333
 QY 841 cacctggaatctgggtgtccctccagtggtggtggttcttcttcttcttcttcttcttct 900
 Db 332 CACCTGGAATNTGGCTGTCCCTCCAGTGGCCATCTTCTTTGTTATCTACACTGACCAG 273
 QY 901 gctggacagtggaatacacagtgtgtcccaagagagcccaactca-ttccaaagccgct 959
 Db 272 GCTGGACAGTGGCGGAATACAGTGTGTGCCCAAGGAGGCCCACTCATTTCAAAAGCCGGCT 213
 QY 960 gccctgccagagccatggcggtgtcttcggagagccctggaccaggtcagtggtggt 1019
 Db 212 GCCCTGCCAGAGCCATGGCGGGGTCTTCGGGACGAGGCCCTGACAGGTGAGTGGAT 153
 QY 1020 ccttggtgcatcttcgtccatgcaagcgcttcttcattggtggtggtggtggtggtggtggt 1079
 Db 152 CCCTGGTGTGATCTTCCTCCATGCAAGCGGCTTCATTGGCGGTCCACGACCCGAGAGGG 93
 QY 1080 tgccttgagatggccgtccaccttggccagcgctcaccacgctcaccacacacacacacac 1139
 Db 92 TGCCTTGAGATGGCCCGTGGCCACCTTGGCCGAGCGGTTCATTGGCGGTCCACGACCCGAGAGGG 33
 QY 1140 gtctaataaaaccttccatctcaaa 1164
 Db 32 GTCTAATAAAACCTTCCATCTCATA 8

RESULT 3
 AI760168/c
 LOCUS
 DEFINITION
 wp58f03.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:2369309 3' similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL
 PROTEIN Y6 LIKE ;, mRNA sequence.

ACCESSION
 AI760168
 VERSION
 AI760168.1 GI:5175835
 KEYWORDS
 EST.
 SOURCE
 human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 610)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On May 18, 1998 this sequence version replaced gi:3137149.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

EMAIL: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 704 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 451.

```

FEATURES
  source
    Location/Qualifiers
      1. .610
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_image="2369309"
        /clone_lib="Soares NSF_F8_9W_OT_PA_P_S1"
        /lab_host="DH10B"
        /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
      BASE COUNT      124 a 166 c 185 g 135 t
      ORIGIN
        Query Match      51.6%; Score 605.2; DB 51; Length 610;
        Best Local Similarity 99.5%; Pred. No. 7.9e-149;
        Matches 607; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 559 gaaggagcctcgatcactgaccactaccctgagtcgacgagtgctgacttaatt 618
      |||||||
Db 610 GAGGGAGCCTCGATATGACGTGACACTACCTGAGTGACGAGTGCTGACTTAAT 551
      |||||||

QY 619 cctacctgaaccaccccccacaaagacactgaggcaggttcaagcgtgcaatggtctg 678
      |||||||
Db 550 CCTACCTGGAACACCCCGACCAAGACACTGAGCGAGGTTCAAGCGTCAATGGATCTG 491

QY 679 gtcaagagaggtttctgagagattagatttctaccacacagctggctgcagcccg 738
      |||||||
Db 490 GTTCAAGAGAGGTTTCTGCAGAGATTAGATTTCTACCAACACAGCTGGCTGCCAGCCCG 431

QY 739 gccttggtgaagagcccttgcccagcattccagtgagccaaagtggagagattgtg 798
      |||||||
Db 430 GCCTTGTTGGAAGGCCCTTGCCAGCGATTCCAGTGGACCCAGTGGAGAGATTGTG 371

QY 799 gaactggcgaaggtgcatgtccctggaaggagcactctctaccacctggaatctgggctg 858
      |||||||
Db 370 GAACCTGGCAAGGTGTCATGCCCTGGAAGGAGCATCTCTACCACCTGGAATCTGGGCTG 311

QY 859 tcccctcagtgccattcttctgttatctactgaccagcagctgagcagtgccgaata 918
      |||||||
Db 310 TCCCTCCAGTGGCCATCTCTTTGTATTATCTACATGACCGAGCTGGACAGTGGCGAATA 251

QY 919 cagtggtgccaaaggagcccccactcattccaaagccggtgcccctgccagagccatgg 978
      |||||||
Db 250 CAGTGTGTCGCAAGAGGCCCTACTATTCCAAAGCCGGCTGCCCTGCCAGAGCCATGG 191

QY 979 cggggttctcggagcagggccctggaccaggtcagtggggattccctggctcattctgcgc 1038
      |||||||
Db 190 CGGGGTCTTCGGGACGAGGGCCCTGGACAGGTGAGTGGGATTCCTGGCTGCATCTTCGTC 131

QY 1039 catcgaagcgtctcattggcgggtccacgcagcagaggggtgctctgagcagtgcccg 1098
      |||||||
Db 130 CATGCAAGCGGCTTCACTGGGGTCAACACACCGAGAGGGTGCTTGAGCATGGCCCGT 71

QY 1099 gccacctggcccgcgtcatacctcccacaaatctcctagcttaataaaaccttccat 1158
      |||||||
Db 70 GCCACCTTGGCCGCGCTATACCTCCCGCAATCTCCTAGTCTAATAAAACCTTCCAT 11

QY 1159 ctcaaaaaa 1168
      |||||||

```

```

Db 10 CTCAAAAAA 1

RESULT 4
AW025001/c
LOCUS
DEFINITION
  AW025001 609 bp mRNA EST 27-OCT-1999
  similar to WP:X08110.8 CE18880 YEAST HYPOTHETICAL PROTEIN YEF6 LIKE
  ; mRNA sequence.
ACCESSION
  AW025001 GI:5878531
VERSION
  AW025001.1
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 609)
  NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  On Jun 5, 1998 this sequence version replaced gi:3187082.
  Contact: Robert Strausberg, Ph.D.
  Tel.: (301) 496-1550
  Email: Robert_Strausberg@nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: M. Bento Soares, Ph.D.
  cDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone distribution: NCI-CCAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -400p from Gibco
High quality sequence stop: 450.
Location/Qualifiers
  1. .609
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_image="2525431"
    /clone_lib="NCI-CCAP_Kid3"
    /lab_host="DH10B"
    /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
  BASE COUNT      125 a 166 c 181 g 137 t
  ORIGIN
    Query Match      51.4%; Score 602.6; DB 64; Length 609;
    Best Local Similarity 99.3%; Pred. No. 3.8e-148;
    Matches 605; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 562 ggggagcctcgatgcactgaccctaccctgagtcagcaggttgcgcgacttaacct 621
      |||||||
Db 609 GGGGAGCCTCGATATGCATGCACACACCTGACGACGAGTTGCTAGACTTAATCCT 550

QY 622 acctggaacaccccgacaaagacactgagcaggggttcaagcgtgcgaatggatcgtt 681
      |||||||
Db 549 ACCTGGAACACCCCGACCAAGACACTGAGCGAGGTTTCAAGCGTGAATGGATCGTT 490

QY 682 caagaggagttctcagagattattctaccacagctgctgcagcccgcc 741
      |||||||
Db 489 CAAGAGGAGTTTCTGCAGAGATTAGATTCTTCAACACAGCTGGCTGCCAGCCGGGCC 430

QY 742 ttggtggaagagcccttgcacagcagattccagggtggagcccaagtggaagatttg 801
      |||||||

```

|||||
Db 429 TTGGTGAAGAGCCCTTGCCAGCGATTCCAGGTGGACCCAAAGTGGAGAGATTGGAA 370
QY 802 ctggcgaaggtgcatgtccctggaaggagcatctctacacatggaatctggcgtgtcc 861
Db 369 CTGGCGAAAGGTGCATGTCCCTGGAAGGAGCATCTCTACCACTGGAAATCTGGGCTGTCC 310
QY 862 cctccagtgccatctcttggttatctacactgaccaggtgacagtgagcgaatacacg 921
Db 309 CTTCCAGTGGCCATCTTTGTTATCTACATGACAGGCTGGACAGTGGCGAATACAG 250
QY 922 tgtgtgccaaagagccccactcattccaaagccggctgccccctgccagagccatggcgg 981
Db 249 TGTGTGCCAAAGAGCCCACTCATTTCCAAAGCCGGCTGCCCTGCCAGAGCCATGGCGG 190
QY 982 ggtctctggagagagccctggaccaggtgcaatggatccctggtgcattcttgcctcat 1041
Db 189 GGTCTTCGGGACAGGCGCTTGACCAAGGTGAGTGGATCCCTGGCTGCATCTTCGTCCAT 130
QY 1042 gcaagcggcttcattggcgtgacccgacccgagaggtgcttgatgacatggccctgccc 1101
Db 129 GCAAGCGGCTTCAATTGGCGGTACCCGACCCGAGAGGGTGCCTTGAGCATGGCCCGTGCC 70
QY 1102 acctggccagcgctcaccctccacaaatctcctagtcttaataaaaccttccatctc 1161
Db 69 ACCTTGGCCAGCGCTCATACCTCCACAAATCTCCTAGTCTAATAAAACCTTCCATCTC 10
QY 1162 aaaaaaaa 1170
Db 9 ATAAAAAA 1
RESULT 5
AW082806/c
LOCUS
DEFINITION
xc03b01.x1 NCI-CGAP_Co21 Homo sapiens cDNA clone IMAGE:2583145 3'
similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL PROTEIN YEV6 LIKE
; mRNA sequence.
ACCESSION AW082806
VERSION AW082806.1 GI:6037958
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 635)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jul 28, 1997 this sequence version replaced gi:2065398.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 418.
Location/Qualifiers
1. .635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2583145"
/issue_type="NCI-CGAP_Co21"
/lab_host="DH10B"

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Normalized to Cot >500. Average insert size 1.04kb.
Normalized version of NCI-CGAP_Col8. Library constructed
by Life Technologies."
BASE COUNT 132 a 176 c 186 g 139 t 2 others
ORIGIN
Query Match 49.8%; Score 583.6; DB 64; Length 635;
Best Local Similarity 97.8%; Pred. No. 3.8e-143;
Matches 622; Conservative 0; Mismatches 11; Indels 3; Gaps 3;
QY 527 atctgtggacaatggatctcccagtgggcagaggggagcctcgatgacctgacca 586
Db 634 ATGCTGTGACATGGATGTCCTCAAGTGGACAGAGGGGGAGCCTCGATATGCACTGACC- 575
QY 587 ctacctgagtcacagagtgctcgacttaactcactggaaccccccagaccagaca 646
Db 575 CTACCNCTGAGTGCACGAGTTGCTCGACTT-ATCGTACCTGGAA-CACCCGACCCCAAGACA 518
QY 647 ctgagcaggggttcaagcgtgcaatggatctgttcaagaggaggttcttcagagattag 706
Db 517 CTGAGGCAGGGTCCANAGCGTGCATGGATGCTGTTCAAGAGGAGTTTCTGCAGAGATTAG 458
QY 707 attctaccacaacagctgctccagccgcttggtggaagagcccttgcctccagc 766
Db 457 ATTCTACCAACACAGCTGGCTGCCACCCGGCGCTTGGTGGAGAGGCGCTTGCACAGC 398
QY 767 gattccaggtggaccccaagtggagagattgtggaactggcgaaagggtgcatgtccctgga 826
Db 397 GATTCCAGGTGGACCCAAAGTGGAGAGATTGTGGAACCTGGCAAGGTGCATGTCCTTGA 338
QY 827 agagcatctctaccacctggaatctgtggctgttccctccagtgccatcttcttatta 886
Db 337 AGGAGCATCTCTACCACTGGAATCTGGGCTGTCCCTCCAGTGGCCATCTTCTTTGTA 278
QY 887 tctacactgaccaggtgagcagtgccgatacagatgtgtgcccaagagcccaatcat 946
Db 277 TCTACACTGACCAAGGTGGACAGTGGCGAATACAGTGTGTGCCCAAGAGGCCCACTCAT 218
QY 947 tccaaagcggctgcccctgccagagccatggcggggtcttcgggagagagccctggacc 1006
Db 217 TCCAAAGCCGGCTGCCCTGCCAGAGCCATGGCGGGGTCTTCGGGAGAGGCCCTTGACC 158
QY 1007 agtccagtggtatccctggctgcatcttcttcctatcgaagcggttcatttggcggtcacc 1066
Db 157 AGGTCAAGTGGGATCCCTTGGCTGCATCTTCGTCATGCAAGCGGCTTCATTGGCGGTCA 98
QY 1067 gacccgagaggggtgcttgagcagatggcccgctgccaccttggccagcgctcatactcc 1126
Db 97 GCACCCGAGAGGGTGGCTTGGAGCATGGCCGCTGGCACCTTGGCCCGCGCTCATACCTCC 38
QY 1127 cacaatctctagtctataataaaccttccatctca 1162
Db 37 CACAAATCTCTAGTCTAATAAAACCTTCCATCTCA 2
RESULT 6
AW082806/c
LOCUS
DEFINITION
IMAGE:2328102 3' similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL
PROTEIN YEV6 LIKE ; mRNA sequence.
ACCESSION AW082806
VERSION AW082806.1 GI:4988363
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 541)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.


```
QY 805 gcaaaaggtgatcccttggaagagcatctctaccacactggaattggctgtccctc 864
Db 357 GCGAAGGTGCATGTCCTCGAAGGAGCATCTCTACCACCTGGAATYGGCTGTCCCT 298
QY 865 ccagtgccatctcttggttatctaacactgaccaggtgagcagtggaatacagtg 924
Db 297 CCAGTGCCATCTCTTGTGTATCTACACTGACCAAGCTGGACAGTGGCCGAATACAGTG 238
QY 925 gtcccaagagcccaactcattccaaagccgctgcccctgcagagccatgcccgggt 984
Db 237 GTGCCCAAGAGCCCACTCATTTCCAAAGCCGGCTGCCCTGGCAGAGCATGGCGGGT 178
QY 985 cticggagcagggccctggaccaggtcagtggtatccctggctgcatcttcgtccatgca 1044
Db 177 CTTCGGGACGAGCCCTGGACCAAGTCAGTGGGATCCCTGGCTGCATCTTTCGTCCATGCA 118
QY 1045 agcggctcattggcgggtcaccgcacccgagaggggtgcttgagcattggccctggcacc 1104
Db 117 AGCGGCTTCATTGGCGGTCAACCGCACCGAGAGGGTGCCCTTGAGCATGGCCGTCGCCACC 58
QY 1105 ttggccagcgtcatatccctccacaaatctctagtcttaataaaaccttccatctc 1161
Db 57 TTGGCCAGCGCTCATACCTCCCAAAATCTCTAGTCTAATAAACTTCCATCTC 1
```

```
RESULT 8
AI752742/c 555 bp mRNA EST 22-JUN-1999
LOCUS
DEFINITION
cln18e03.x1 Normal Human Trabecular Bone Cells Homo sapiens cDNA
cln18e03.x1 random, mRNA sequence.
AI752742
AI752742.1 GI:5131006
EST.
KEYWORDS
SOURCE
ORGANISM
human.
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.,
Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M.,
Robey,P.G., Hotchkiss,R.N. and Francomano,C.A.
SCAP: The Skeletal Genome Anatomy Project
Unpublished (1997)
JOURNAL
On May 9, 1996 this sequence version replaced gi:1132592.
COMMENT
Contact: Libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-402-4877
Fax: 301-496-7157
Email: libin@helix.nih.gov
DNA Sequencing and analyses by National Institutes of Health
Intramural Sequencing Center (NISC).
Plate: 18 row: e column: 03
Seq primer: -21m13 forward primer (ABI).
FEATURES
source
1. .555
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NHTBC_cln18e03"
/clone_lib="Normal Human Trabecular Bone Cells"
/sex="Female"
/tissue_type="Bone"
/cell_type="Trabecular Bone Cells"
/lab_host="SURE"
/note="Organ: Hip; Vector: pBluescript; Site_1: EcoRI"
```

```
BASE COUNT 112 a 147 c 162 g 134 t
ORIGIN
```

Query Match 45.1%; Score 529.6; DB 51; Length 555;
Best Local Similarity 99.1%; Pred. No. 5.7e-129;
Matches 543; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
QY 627 gaaccaccgacccaagacactaagcaggttcaacgtgcaatgattcgttgc-aag 685
Db 555 GACCACCCCGACCAAGACACTGAGCGAGGTTTCAAGCGTCAATGATTGGTTCAAAA 496
QY 686 aggaagtctgcagagatttagatttaccacaacacagctggtgccagccggccttgg 745
Db 495 AGAAGTTTCTGCAGAGATTAGATTTCACCAACACACAGCTGGCTGCCAGCCCGGCGCTGG 436
QY 746 tgaagaggcccttgcccagcagattccaggtgagcccaagtggagagattgtgaaatgg 805
Db 435 TGAAGAGGCGCCTTGCCAGCGATTCCAGGTGGAGCCCAAGTGGAGAGATTGTGGAACGTG 376
QY 806 cgaagtgcattccctggaagagcatctaccacactggaatctggctgtccctc 865
Db 375 CGAAGGTGTCATGTCCCTGGAAGGAGCATCTTACCACCTGGAATCTGGGCTGTCCCTC 316
QY 866 cagtggcatctcttctgttatctacactgaccaggtggacagtgcgcaatacagtg 925
Db 315 CAGTGGCCATCTTCTTGTATCTACACTGACCAAGCTGGACAGTGGGGAATACAGTGTG 256
QY 926 tgcccaaggagcccaactcattccaaagccggtgcccctgccaagccatggcggggtc 985
Db 255 TGCCCAAGGAGCCCACTCATTTCCAAAGCCGGCTGCCCTGCCAGAGCCATGGCGGGTC 196
QY 986 ttcgggagcagggccctggaccaggtcagtgggatccctggctgcatcttgcctcatcaa 1045
Db 195 TTGGGACGAGGCGCCTGGACAGTCAAGTGGGATCCCTGGCTGCATCTTGTTCATGCAA 136
QY 1046 gcggctcattggcgggtcacgcacccgagaggtgcttgagcattggccgtgccacct 1105
Db 135 GCGGCTTCATTGGGGTCCACCGCACCGAGAGGGTGGCTTGAGCATGGCCGTCGCACCT 76
QY 1106 tggccagcgtcatacctcccaaaatctctagtcttaataaaaccttccatctcaaaa 1165
Db 75 TGGCCCAAGCGCTCATACCTCCCAAAATCTCTAGTCTAATAAAACCTTCCATCTCTATA 16
QY 1166 aaaaaaa 1173
Db 15 AAAAAAA 8
```

RESULT 9
AI936652/c 536 bp mRNA EST 06-SEP-1999
LOCUS
DEFINITION
wp68b08.x1 NCI-CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2466903 3'
similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL PROTEIN YEY6 LIKE
/, mRNA sequence.
AI936652
AI936652.1 GI:5675533
EST.
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 536)
NCI/MINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
TITLE
(CGAP/BTGP), Tumor Gene Index
JOURNAL
Unpublished (1998)
COMMENT
On Jun 5, 1998 this sequence version replaced gi:3187056.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

Insert Length: 1068 Std Error: 0.00
Seq primer: -40ml3 fwd. EV from Amersham
High quality sequence stop: 444.
Location/Qualifiers
1. 492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1737522"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTACCAATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 101 a 133 c 150 g 108 t
ORIGIN
Query Match 40.3%; Score 473.2; DB 42; Length 492;
Best Local Similarity 99.2%; Pred. No. 3.6e-114;
Matches 486; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 685 gagaggtttctgcagagattagattttctacaaacagctggctgcagccgggcttg 744
DB 492 GAGGAGTTTCTGAGAGATTAGATTCTTACCAACACAGCTGGCTGCCAGCGGCGCTTG 433
QY 745 gtggaagagccctgccagcgattccaggaggagcccaagtggaagattgtggaactg 804
DB 432 GTGGAAGAGCGCCCTTCCAGCGATTCCAGGTGGACCCCAAGTGGAGAGATTGTGGAACGTG 373
QY 805 gcgaaggtgcattgtccctgggaagagagatctctaccacccctggaatctgagctgtccctc 864
DB 372 GCGAAGGTGCATGTCCCTGGAAGGAGCATCTCTACCACTGGGAATCTGGGCTGTCCCT 313
QY 865 ccagtgccattctttgtttattctacactgacagctggagctggagctggagctggagctgg 924
DB 312 CCAGTGGCCATCTCTTTGTTATCTACACTGACAGGCTGGAGCTGGAGCTGGAGCTGGAG 253
QY 925 gtgccaagagagccctcattccaaagccggtgctccctgccagagccatgagcagctgagc 984
DB 252 GTGCCCCAAGGAGCGCCCATCTATTCCAAAGCGCTGCCCTGCCAGAGCATGGCGGGGT 193
QY 985 cttcgggagagccctggaccaggtcagtggaatccctggctgcatcttcgtccatgca 1044
DB 192 CTTGGGAGAGCGCCCTGGACAGGTGAGTGGGATCCCTGGCTGCTGCTGCTGCTGCTGCTG 133
QY 1045 agcggtctcattgagc-gtcacccgacccgagaggtgctgctgagcagctgagcagctgagc 1103
DB 132 AGCGGTTCTATTGGCGCGCCCGCCGACCCGAGAGGGGTGCTTGGAGCTGGCGCTGCCAC 73
QY 1104 cttgccccagcgtcctatacctcccaaaatctctctagctctataataaaaccttcacataa 1163
DB 72 CTTGGCGCAGCGCTCATACCTCCCAAAATCTCTAGTCTAATAAAACCTTCATCTCAT 13
QY 1164 aaaaaaaaaa 1173
DB 12 AAAAAAAAAA 3

FEATURES
source
1. 492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1737522"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTACCAATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 101 a 133 c 150 g 108 t
ORIGIN
Query Match 40.3%; Score 473.2; DB 42; Length 492;
Best Local Similarity 99.2%; Pred. No. 3.6e-114;
Matches 486; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 685 gagaggtttctgcagagattagattttctacaaacagctggctgcagccgggcttg 744
DB 492 GAGGAGTTTCTGAGAGATTAGATTCTTACCAACACAGCTGGCTGCCAGCGGCGCTTG 433
QY 745 gtggaagagccctgccagcgattccaggaggagcccaagtggaagattgtggaactg 804
DB 432 GTGGAAGAGCGCCCTTCCAGCGATTCCAGGTGGACCCCAAGTGGAGAGATTGTGGAACGTG 373
QY 805 gcgaaggtgcattgtccctgggaagagagatctctaccacccctggaatctgagctgtccctc 864
DB 372 GCGAAGGTGCATGTCCCTGGAAGGAGCATCTCTACCACTGGGAATCTGGGCTGTCCCT 313
QY 865 ccagtgccattctttgtttattctacactgacagctggagctggagctggagctggagctgg 924
DB 312 CCAGTGGCCATCTCTTTGTTATCTACACTGACAGGCTGGAGCTGGAGCTGGAGCTGGAG 253
QY 925 gtgccaagagagccctcattccaaagccggtgctccctgccagagccatgagcagctgagc 984
DB 252 GTGCCCCAAGGAGCGCCCATCTATTCCAAAGCGCTGCCCTGCCAGAGCATGGCGGGGT 193
QY 985 cttcgggagagccctggaccaggtcagtggaatccctggctgcatcttcgtccatgca 1044
DB 192 CTTGGGAGAGCGCCCTGGACAGGTGAGTGGGATCCCTGGCTGCTGCTGCTGCTGCTGCTG 133
QY 1045 agcggtctcattgagc-gtcacccgacccgagaggtgctgctgagcagctgagcagctgagc 1103
DB 132 AGCGGTTCTATTGGCGCGCCCGCCGACCCGAGAGGGGTGCTTGGAGCTGGCGCTGCCAC 73
QY 1104 cttgccccagcgtcctatacctcccaaaatctctctagctctataataaaaccttcacataa 1163
DB 72 CTTGGCGCAGCGCTCATACCTCCCAAAATCTCTAGTCTAATAAAACCTTCATCTCAT 13
QY 1164 aaaaaaaaaa 1173
DB 12 AAAAAAAAAA 3

RESULT 14
AA18018/c
LOCUS
DEFINITION
AA18018 600 bp mRNA EST 03-JUL-1999
UI-R-A0-ah-d-10-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
UI-R-A0-ah-d-10-0-UI 3' similar to spIP400931YEY6_YEAST
HYPOTHETICAL 38.2 KD PROTEIN IN BEM2-NCB1 INTERGENIC REGION

consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 101 a 132 c 147 g 97 t
ORIGIN
Query Match 40.5%; Score 475.4; DB 48; Length 477;
Best Local Similarity 99.8%; Pred. No. 9.4e-115;
Matches 476; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 685 gagaggtttctgcagagattagattttctacaaacagctggctgcagccgggcttg 744
DB 477 GAGAGGTTTCTGAGAGATTAGATTCTTACCAACACAGCTGGCTGCCAGCGGCGCTTG 418
QY 745 gtggaagagcccttgcagcgattccaggaggagcccaagtggaagattgtggaactg 804
DB 417 GTGGAAGAGCGCCCTTGCCAGCGATTCCAGGTGGAGCCCAAGTGGAGAGATTGTGGAACGTG 358
QY 805 gcgaaggtgcattgtccctgggaagagcattctaccacctggaatctgagctgtccctc 864
DB 357 GCGAAGGTGCATGTCCCTGGAAGGAGCATCTCTACCACTGGAAATCTGGGCTGTCCCT 298
QY 865 ccagtgccattctttgtttattctacactgacagcgctggagctggagctggagctggagctgg 924
DB 297 CCAGTGGCCATCTCTTTGTTATCTACACTGACAGGCTGGACAGTGGCGAATACAGTGT 238
QY 925 gtgccaagagagccctcattccaaagccggtgctccctgccagagccatgagcagctgagc 984
DB 237 GTGCCCCAAGGAGCGCCCATCTATTCCAAAGCGCTGCCCTGCCAGAGCATGGCGGGGT 178
QY 985 cttcgggagagccctggaccaggtcagtggaatccctggctgcatcttcgtccatgca 1044
DB 177 CTTGGGAGAGCGCCCTGGACAGGTGAGTGGGATCCCTGGCTGCTGCTGCTGCTGCTGCTG 118
QY 1045 agcggtctcattgagc-gtcacccgacccgagaggtgcttggagctggagctggagctggagctgg 1104
DB 117 AGCGGTTCTATTGGCGGTGACCGCACCGAGAGGGTGCCTTGAGCATGGCGCGCTGGCAC 58
QY 1105 ttgccccagcgtcattccctcccaaatctctagctctataataaaaccttcacatc 1161
DB 57 TTGGCCAGCGCTCATACCTCCCAAAATCTCTAGTCTAATAAAACCTTCATCTC 1
RESULT 13
AA18745/c
LOCUS
DEFINITION
AA18745 492 bp mRNA EST 28-OCT-1998
qd98c10.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1737522
3' similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL PROTEIN YEY6
LIKE ; mRNA sequence.
ACCESSION AA138745
VERSION AA138745.1 GI:3644717
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 492)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2285431.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

gi|1107718|pir||S50659 hypothetical protein YER156c - yeast
[Saccharomyces cerevisiae] gi|603396 (U18917) YER156cp
AA818018
AA818018.1 GI:4198541
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 600)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
On Feb 17, 1998 this sequence version replaced gi:2887898.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult spleen library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics This clone is also available through the I.M.A.G.E.
Consortium at LNL (info@image.llnl.gov). IMAGE ID=1767406
Seq primer: M13 Forward
POLYA=No.

FEATURES

Location/Qualifiers

1..600
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-ah-d-10-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture."

BASE COUNT 121 a 161 c 172 g 146 t

ORIGIN

Query Match 37.7%; Score 442.8; DB 45; Length 600;
Best Local Similarity 86.5%; Pred. No. 3.9e-106;
Matches 514; Conservative 0; Mismatches 72; Indels 8; Gaps 2;
Qy 588 taccctgagtcacgagttgtcgtacctaatctacccggaaccccccagaccagacac 647
Db 600 TACACTGAGTCCCGGGTTTCTCGCTTAATCCCACTGGAACCAAGCCGACCAAGACAC 541
Qy 648 tgaggcagggttcgaagctgcaatggatctggttcaagagagattctgcagagattaga 707
Db 540 TGAGGACAGGATTACAGCGCTGGCGATGACCTGTGCGAAGAGAGTTTCGCAAGACTAAA 481
Qy 708 ttctaccacaacagctggtgcagcccgcccttgtgaaagagccctgtgccagcg 767
Db 480 CTTCTACCAAGACACTGCTGCTCCAGCCCGGCCCTTGTAGAGAGAGCCCTGGCCAGAG 421
Qy 768 attccagtggaaccaaagtggagagattgtgaaactgagcgaagatgcatgtccctggaa 827
Db 420 ATTCAAGGTAGACTCACTGAGGGAGATAGTGAACCTTCCAAAGGCGGATCCCTTTGAA 361

Qy 828 ggagcatctaccacotgggaatctgggctgtccctccagtgcccatcttctttgtat 887
Db 360 CGAGCACTGTACCACCTAGAAATCTGACTGTCTCCACAGTAGCCATCACCCTTTGTAT 301
Qy 888 ctacactgaccagctgacagtgccgaataacagtggtgtgcccaggagcccccatt 947
Db 300 CTACACTGACCAAGCTGGACAGTGGCGAGTCCAGTGTGTGCCAAGAGGCCTCACTCAT 241
Qy 948 ccaagccggctgccccctgcagagccaatgagggtttctgggacgagggccctggagaca 1007
Db 240 CCAAGCCGGCTGCCCTGCCGAGCCATGGCGGACATTCGGGATGAGGCCCTGGAGCA 181
Qy 1008 ggtcagtgaggtccctggtcgtcatcttgcgtccatgcaagcggtcttcattggcggtcacgc 1067
Db 180 AGTCAGTGGGATCCCTGGTTCATCTTCGTCATGCCAGTGGCTTCATCTGGTGGCACA 121
Qy 1068 caccgagaggggtgcttgagcatggccctggccaccttgcccgagcgctca-----t 1120
Db 120 CACTCGAGAGGGTGGCCCTGAACATGGCCCGTGCCACCTTGCCAGCGCACAGCACCTGT 61
Qy 1121 acctccacaatctcttagtctataaaacct-tccatctctcaaaaaa 1173
Db 60 ACCTCTTGCAAAATGCTGTAGTCCCAATAAACTTATTTCATCTCTATAAAAAA 7

RESULT 15

AI277497/c

LOCUS

DEFINITION

gi96d08.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1880175
3' similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL PROTEIN YEF6
LIKE ; mRNA sequence.

ACCESSION

AI277497

VERSION

AI277497.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Aug 26, 1998 this sequence version replaced gi:3476385.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1053 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 362.

FEATURES

source

1..435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1880175"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbM, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

Wed May 24 09:19:45 2000

us-09-215-435-117.rst

```
BASE COUNT      90 a      122 c      136 g      87 t
ORIGIN

Query Match      36.7%; Score 430.8; DB 44; Length 435;
Best Local Similarity 99.5%; Pred. No. 4.8e-103;
Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 731 cagccgggcttggtggaagagagcccttccagcgattccaggtggacccaagtggag 790
Db 434 CAGCCCGGGCTTGGTGGAGAGAGCCCTTCCCAAGCGATTCCAGGTGGACCAAGTGGAG 375
QY 791 agatttggaactggcgaaagggtgcattccctgggaaggagcatctctaccacctggaat 850
Db 374 AGATTGTGGAAC TGGCAAGAGGTGCATGTCCCTGGAAGGAGCATCTCTACCACTGGAAT 315
QY 851 ctggggtgtccctccagtggtggaatctctttgttattctacactgaccaggtggacagt 910
Db 314 CTGGGCTGTCCCTCCAGTGGCCATCTCTTTGTATTCTACTGACCAAGCTGGACAGT 255
QY 911 ggcgaatacagtggtgcccgaagagccccactcattccaaagccggctgccccctgccag 970
Db 254 GCGCAATACACTGTGTGCCCAAGGAGCCCACTCATTTCCAAAGCCGCTGCCCTGCCAG 195
QY 971 agccatggcgggtcttctgggacagagccctggaccaggtcagtgggatccctggctgca 1030
Db 194 AGCCATGGCGGGGTCTTCGGGACAGAGCCCTTGGACCAAGGTCACTGGGATCCCTGGCTGCA 135
QY 1031 tcttcgtccatgcaagcggtctcattggcggtcaccgcacccgagaggtgcttgagca 1090
Db 134 TCTTCGTCCATGCAAGCGGCTTCATTGGCGGTCAACCGCACCCGAGAGGGTGCCTTGAGCA 75
QY 1091 tggcccggtgccaccttggcccgagcgtcatatccctccacaaaatctctagtctataaaaa 1150
Db 74 TGGCCCGTGGCCACCTTGGCCAGCGCTCATACCTCCCAAAATCTCTAGTCTATAAAA 15
QY 1151 ccttcacatcctcaaa 1164
Db 14 CCTTCCATCTCATATA 1
```

Search completed: May 23, 2000, 22:03:49
Job time: 3987 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:38:17 ; Search time 6115.43 Seconds
(without alignments)
-124.871 Million cell updates/sec

Title: US-09-215-435-118
Perfect score: 785
Sequence: 1 cggatccggaggtccggtg.....tatggaaataaagttctttc 785

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:**
1: gb_ba1:**
2: gb_ba2:**
3: gb_cm:**
4: gb_ov:**
5: gb_pat:**
6: gb_ph:**
7: gb_pl1:**
8: gb_pl2:**
9: gb_pr1:**
10: gb_pr2:**
11: gb_pr3:**
12: gb_ro:**
13: gb_sts:**
14: gb_sy:**
15: gb_un:**
16: gb_v1:**
17: em_fun:**
18: em_hum1:**
19: em_hum2:**
20: em_in:**
21: em_om:**
22: em_or:**
23: em_ov:**
24: em_pat:**
25: em_ph:**
26: em_pl:**
27: em_ro:**
28: em_sts:**
29: em_sy:**
30: em_un:**
31: em_v1:**
32: gb_htg1:**
33: gb_htg2:**
34: gb_in1:**
35: gb_in2:**
36: em_ba1:**
37: em_ba2:**
38: em_hum3:**
39: em_hum4:**
40: gb_pr4:**
41: gb_htg3:**
42: gb_htg4:**
43: gb_htg5:**
44: gb_htg6:**

45: gb_htg7:**
46: em_htg1:**
47: em_htg2:**
48: em_htg3:**
49: em_hum5:**
50: gb_pl3:**
51: gb_pr5:**
52: gb_htg8:**
53: gb_htg9:**
54: gb_htg10:**
55: gb_htg11:**
56: gb_htg12:**
57: gb_htg13:**
58: gb_htg14:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	660	84.1	908	40	AF177398 Homo sapi
2	527.4	67.2	216002	33	HSJ854E16
C 3	392.4	50.0	38805	54	AC010643 Homo sapi
4	239.4	30.5	828	12	AF177399 Mus muscu
C 5	177.8	22.6	82585	45	AC010524 Homo sapi
6	81	10.3	82585	45	AC010524
7	54.8	7.0	5364	1	SGKSACPG
8	45.6	5.8	7877	3	BTBLGG
9	45.6	5.8	9432	3	BTBLGLOBB
C 10	45	5.7	58842	42	AC015485
C 11	43.8	5.6	23536	34	CEF57B7
12	43	5.5	7218	5	I66494
13	42.4	5.4	32039	1	SC7A1
14	42	5.4	3768	12	MMP4G
15	42	5.4	6706	9	AB018254
16	42	5.4	130831	41	AF170801
17	42	5.4	200945	45	AC019257
18	41.4	5.3	897	12	MMRP28
19	41.4	5.3	1170	12	MMRP4628
20	41.4	5.3	4936	1	ECR751TRA
21	41.4	5.3	34074	1	SCF81
C 22	41.4	5.3	53339	2	EAU67194
23	41.2	5.2	5659	9	AB020628
24	41	5.2	699	5	I09028
25	41	5.2	1134	5	I03259
26	41	5.2	1134	5	I09027
27	41	5.2	1895	34	PFACSA
28	41	5.2	1908	5	I08489
29	41	5.2	6461	2	AF204400
30	40.6	5.2	29688	34	CEC09G5
31	40.4	5.1	2817	34	AF015539
32	40.4	5.1	11442	34	CEH06A10
C 33	40.4	5.1	24700	1	SCE8
34	40.2	5.1	18302	34	CEY2H9A
35	40.2	5.1	22449	1	SPSNECDE
36	40.2	5.1	22449	1	SPSNBCGEN
C 37	40.2	5.1	33779	1	SCGD3
C 38	40.2	5.1	41322	34	CEC50B6
39	39.6	5.0	3428	9	AB011174
40	39.6	5.0	3780	1	AB024740
41	39.6	5.0	145839	42	AC011719
C 42	39.6	5.0	162114	16	AF157706
43	39.4	5.0	479	13	G33950
44	39.4	5.0	1278	34	PFACSAAD
45	39.4	5.0	3072	8	PCAF001305

ALIGNMENTS

Wed May 24 09:19:45 2000

```

RESULT 1
AF177398 908 bp mRNA PRI 16-OCT-1999
LOCUS Homo sapiens soggy-1 protein (SGY-1) mRNA, complete cds.
DEFINITION AF177398
ACCESSION AF177398.1 GI:6049611
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 908)
AUTHORS Krupnik,V.E., Sharp,J.D., Jiang,C., Robison,K., Chickering,T.W.,
Amaravadi,L., Brown,D.E., Guyot,D., Mays,G., Leiby,K., Chang,B.,
Duong,T., Goodearl,A.D.J., Gearing,D.P., Sokol,S.Y. and
McCarthy,S.A.
TITLE Functional and structural diversity of the human Dickkopf gene
family
JOURNAL Gene 238 (2), 301-313 (1999)
MEDLINE 20035735
REFERENCE 2 (bases 1 to 908)
AUTHORS Krupnik,V.E., Sharp,J.D., Jiang,C., Robison,K., Chickering,T.W.,
Amaravadi,L., Brown,D.E., Guyot,D., Mays,G., Leiby,K., Chang,B.,
Duong,T., Goodearl,A.D.J., Gearing,D.P., Sokol,S.Y. and
McCarthy,S.A.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Cell Biology, Millennium Biotherapeutics
Inc., 640 Memorial Drive, Cambridge, MA 02139, USA
FEATURES
source
location/Qualifiers
1..908
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..908
/gene="SGY-1"
66..794
/gene="SGY-1"
/translation="MGASPPAPARRHLVLLLLSTLVIPSAAPAIHDADAQESSLG
LTLQSLQSFRLFLKGNLLRGIDSLFSAFMDRFLGPNYHKEENOHQNLNTLS
HLQDKMTDNKTEVLISENVASIQAESFEGDLKVPMEKEALVPIQKATDSFH
TELPRVAFWIKLPRRSHDQALEGGHMLSEKRLQLAIRDGLRGTGTHKDVLEETE
SSSHSRISPRKTHLYILRPSROL"
BASE COUNT 228 a 283 c 249 g 148 t
ORIGIN
Query Match 84.1%; Score 660; DB 40; Length 908;
Best Local Similarity 89.1%; Pred. No. 3.5e-125;
Matches 763; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 23 ccgggtgtggtctagcataaaggcgagccagagagggcggtgtggtatggagaagc 82
Db 17 CCGGGTGTGTTAGCATAAAGCGGAGCCAGAGAGGGCGGGTATGGGAGAGC 76
QY 83 ctccacacctgcccgcgaaggcggtcatctgtgtctgtctgtctgtctctctacct 142
Db 77 CTCCCCACCTGCCCCGCAAGGCGGCATCTGCTGCTGCTGCTGCTGCTGCTGCT 136
QY 143 ggtgataccctcgcgtgcagctctctctctctctctctctctctctctctctctct 202
Db 137 GGTGATCCCTCCGCGTCAGCTCTTATCATGATGCTGACGCCCAAGAGAGCTCTTGG 196
QY 203 tctcaggcctccagagcctactccaaggcttcagccgacctttctgaaaggtaacct 262
Db 197 TCITCAGGCTCCAGAGCCTACTCCAAGGCTTCAGCCGACITTTCTGAAAGGTAAC 256
QY 263 gcttcggggcatagacagcttattctctcccccagcttcggggcgctccctgggaa 322

```

```

Db 257 GCTTCGGGCATAGACAGCTTATTCTCTGCCCCCATGGAGCTTCGGGGGCGCTCCCTGGGAA 316
QY 323 ctaccacaaggaggagaaaccagagaccagctgggggaaacacacacctctccagccact 382
Db 317 CTACCACAAGAGGAGAGACAGGAGCAGCTGGGGGAAACACACCTCTCCAGCACCT 376
QY 383 ccagatcgac----- 393
Db 377 CCAGATCGACAAGATACCGACGAACAAGACAGGAGAGGTGCTGATCTCCGAGAGATGTGT 436
QY 393 -----aaggtaccaccaggatga 409
Db 437 GGCATCCATTCAACACGAGGGAGCTTCGAGGGTGATTTGAAGGTACCCAGGATGGA 496
QY 410 ggagaaggagggcctgtgtacccatccagaaagccagagcagcttcacacagaactcca 469
Db 497 GGAGAAGGAGGCGCTGTGTACCCATCCAGAAAGCCAGCGACAGCTTCCACACAGAACTCCA 556
QY 470 tccccgggtgccttctgtgattcattaaagctgccacgagggaggtcccccagagatccct 529
Db 557 TCCCCGGGTGGCTTCTGGATCATTAAGCTGCCACGGGAGGTCCACACAGGATGCCCT 616
QY 530 ggaggggcgccactggtctcagcagagagcagacccgctcagccatccgggatggact 589
Db 617 GGAGGGCGGCACCTGGCTCAGCGAGAAAGCGACACCGCTCGAGGCCATCCGGGATGGACT 676
QY 590 ccgaaggggaccacaaaggagcgtctcagaaagggagcagagagctctcccccactccag 649
Db 677 CCGCANGGGGACCCACACAGGAGCTCTTAGAAGAGGGGACCGAGAGCTCCTCCCACTCCAG 736
QY 650 gctgtcccccggaaagacccacttactgtactactcagggccctcagcagctctgaggg 709
Db 737 GCTGTCCCGCCGGAAGACCCACTTACTGTACATCTCAGGGCCCTCTCGGACGCTGTAGGG 796
QY 710 gtggggagcggggagcactcctctgtagcccccacagacctgcccccaagcaccatgt 769
Db 797 GTGGGGACCGGGGAGCACCTGCTGTAGCCCCCATCAGACCCCTGCCCAACGACCATATG 856
QY 770 gaataaaagtctcttc 785
Db 857 GAAATAAGTTCTTTC 872
RESULT 2
HSJ854E16 216002 bp DNA HTG 08-FEB-2000
LOCUS Homo sapiens chromosome 20 clone RP5-854E16, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL121723
VERSION AL121723.26 GI:6982994
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 216002)
AUTHORS Blakey,S.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Feb 16, 2000 this sequence is unfinished and does not necessarily
requests: clonerequest@sanger.ac.uk
IMPORTANT: This sequence is unfinished. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00622 Length: 1746bp
Contig_ID: 00669 Length: 46937bp
Contig_ID: 02054 Length: 165719bp.
* NOTE: This is a 'working draft' sequence. It currently
consists of 3 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1746: contig of 1746 bp in length
* 1747 2546: gap of 800 bp
* 2547 49483: contig of 46937 bp in length
* 49484 50283: gap of 800 bp
* 50284 216002: contig of 165719 bp in length.

FEATURES

source

1. .216002
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-854E16"
/clone_11b="RPC1-5"
BASE COUNT 69883 a 40683 c 46028 g 57808 t 1600 others
ORIGIN

Query Match 67.2%; Score 527.4; DB 33; Length 216002;
Best Local Similarity 79.9%; Pred. No. 1.4e-98;
Matches 701; Conservative 0; Mismatches 81; Indels 95; Gaps 3;

Qy	1	cgaatccggagtcggtaccggtgctggtctagcataaaggcggagccagaaga	60
Db	61265	CAGAGTCTGGAGTGCAGTGA-CAGGCTGTGTCTAGCTAAGCGGAGCCAGAGA	61323
Qy	61	aggggcgggtatggagaagcctccacactgcccccgcaaggcgcattctgtgtcc	120
Db	61324	AGGGCGTGGGTGGAAGAACTCTCTACTCCCGCGCAATGCGCATCCGCTGTCC	61383
Qy	121	tgctgtgctctctatccctggatccctcgcctcgcctcgcctcctatcatgctg	180
Db	61384	TGCTGTGCTCTCTCTGCGCTGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT	61443
Qy	181	acgcccagaagagctctgtggtctcagagctccagacctactcgaagctcagcc	240
Db	61444	ATGCCAAGAGAGCTCTCTGGTCTTACAGCCCTCCAGAGCTACTCCAGGCTTCAG	61503
Qy	241	gacttttctgaaaggtaacctgtctcgggcatagacagcttattctgtccccatgg	300
Db	61504	GACTTTTCTGAAAGATGACTGCTGTCGGGCATAGACAGCTTCTTCTGCCCCATGG	61563
Qy	301	acttccgggctcctcgggaactaccacaaagaggaggaaccagagcaccagctggga	360
Db	61564	ACTTCCGGGGCTCTCCCTAGGAACCTACCAACAGAGAGAAACGAGGAGCACCATG	61623
Qy	361	acaacacctctccagcaccctccagatcac-----	393
Db	61624	ACAACACCTCTCTCCAGCCACTCTCCATATTGACAGGTGACCGACAAATAGACAG	61683
Qy	393	-----	393
Db	61684	TGCTGATCTCCGAGAAGGTGGTGCATCCATCCAGCCGGCGGAGGGAGCTTCAGGG	61743
Qy	393	-----aaggtaccagagatgagagagagagccttggtaccatccagaagccacgg	447
Db	61744	ACTGGAAGGTACCCAGGATGAGTAGGCGCGGCCCTGGTGTCCATCCGGAAGGCTATGG	61803
Qy	448	acaagttccacagaactccatccctcccggtgtgctctctggatcattaaagctgcacggc	507
Db	61804	ACAACATTCATGACAGAGCTCATCCCGGGTGGCTTCTGGATCATGAAGCTGCCACGGT	61863
Qy	508	ggaggtccaccagatgctcctggagggcgccactgctcagcagagaagcacccgcc	567
Db	61864	GGAGGTCCACACCAATGTCTTGAGGGGGCGCCCTGGCTCAGTGAAGAGGACACCGCC	61923
Qy	568	tgcagggccatccgggatggactccgcaaggggaccaccaaggagcgtctcagaagggga	627
Db	61924	TGCAGGGCCATCCAGGATGGGCTCCACAGAGGGGACCCCGAGGAGCGTCTTAAACAGG	61983

Qy	628	ccgagagctctccactccaggtgtcccccgaagaccacttactgtacatctcca	687
Db	61984	CCCAGGCTCTCCACTCCCGGGCTGT-CTCCGAAAGACCCACTTACTGTACATCTCA	62042
Qy	688	ggcctctcggcagctgtaggggtgggagccgggagcacctgctgtagccccatcag	747
Db	62043	GGCTTCTGTCAGCTATAGGGTTGGACCGGGGAGCACCTGTCTGTAGCCCCCATCAG	62102
Qy	748	acctgtcccaagcaccatatggaataaagtctttt	784
Db	62103	GCCCCGCCCAAGCAGCATATGGAATAAAGTTCTTT	62139

RESULT

3

AC010643/c

LOCUS

DEFINITION

AC010643

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

TITLE

JOURNAL

COMMENT

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 26372: contig of 26372 bp in length

* 26373 28464: contig of 2092 bp in length

* 28465 38805: contig of 10341 bp in length.

FEATURES

source

1. .38805

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/clone="LLNLR-254A7"

BASE COUNT 9218 a 9816 c 9791 g 9830 t 150 others

ORIGIN

Query Match 50.0%; Score 392.4; DB 54; Length 38805;

Best Local Similarity 99.7%; Pred. No. 6.1e-71;

Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 392 caaggtaccagagtgaggagaaggagccctggtaccatccagaagccagacag 451

Db 38410 CCAGGTACCCAGGATGGAGGAGGAGGAGGCCCTGTGTACCCATCCAGAGGCCACGACAG 38351

QY 452 cttccacacagaactccatccccgggtggtcttctggtatcattagctgcacagcgag 511
Db 38350 CTTCCACACAGAATCCATCCCGGGTGGCTTCTGGATCATTAAGCTGCCACGCGGAG 38291
QY 512 gtccacacagatgacctgagggcgccactggtctcagcagagagcagacccgctgca 571
Db 38290 GTCCACACAGATGCCCTGGAGGGGGGGCCACTGGCTACGCGAGAGCAGACCCGCTGCA 38231
QY 572 ggcacccggatgagctccgcaagggggaccacacagagcgtccctagagggggaccga 631
Db 38230 GGCATCCGGATGAGCTCCGCAAGGGGACCCACAGGAGCTCTCTAGAGAGGGGACCGA 38171
QY 632 gagctcctccactcagagctgtcccccgaagagacccacttactgtatcactcagacc 691
Db 38170 GAGCTCTCCACTCAGGCTGTCCCGCCGAAAGACCCACTTACTGTATCCTCAGGCC 38111
QY 692 ctctggcagctaggggtggggagcggggagcactgctgtagccccccatcagacc 751
Db 38110 CTCTGGCACTGTAGGGTGGGACCGGGGAGCAGCTGCTGTAGCCCCCATCAGACCC 38051
QY 752 tgccccagcaccatggaataaagtctttc 785
Db 38050 TGCCCCAAGCACCATATGGAATAAAGTTCTTTC 38017
RESULT 4
AF177399
LOCUS AF177399 828 bp mRNA ROD 16-OCT-1999
DEFINITION Mus musculus soggy-1 protein (Sgy-1) mRNA, complete cds.
ACCESSION AF177399
VERSION AF177399.1 GI:6049613
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 828)
AUTHORS Krupnik,V.E., Sharp,J.D., Jiang,C., Robison,K., Chickering,T.W.,
Anaravadi,L., Brown,D.E., Guyot,D., Mays,G., Leiby,K., Chang,B.,
Duong,T., Goodearl,A.D.J., Gearing,D.P., Sokol,S.Y. and
McCarthy,S.A.
TITLE Functional and structural diversity of the human Dickkopf gene
family
JOURNAL Gene 238 (2), 301-313 (1999)
MEDLINE 20035735
REFERENCE 2 (bases 1 to 828)
AUTHORS Krupnik,V.E., Sharp,J.D., Jiang,C., Robison,K., Chickering,T.W.,
Anaravadi,L., Brown,D.E., Guyot,D., Mays,G., Leiby,K., Chang,B.,
Duong,T., Goodearl,A.D.J., Gearing,D.P., Sokol,S.Y. and
McCarthy,S.A.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Cell Biology, Millennium BioTherapeutics
Inc., 640 Memorial Drive, Cambridge, MA 02139, USA
FEATURES
source Location/Qualifiers
1..828
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..828
/gene="Sgy-1"
CDS 51..743
/gene="Sgy-1"
/note="secreted protein"
/codon_start=1
/product="soggy-1 protein"
/protein_id="AAF02679.1"
/db_xref="GI:6049614"
/translation="MCLRVLILLPLAFVSSNALPHDVSQNTSGFLQLRLLOS
FSRLFLNDLRLDNFSPFRDLPRNFHOENGEMKGNHTLSSHLDIKVTDN
QTGVHTSEKYEASIEPERNPEDMKVPKVEAPPPVOKVDSLHPERQVAFWIM
KMPRRTPQPDVQDGRWLIEKRHRMQAIRDLRGAREDSLEDGVHIPQHKLPVRK
HFLYLRLPSQQL"
BASE COUNT 223 a 250 c 209 g 146 t
ORIGIN

Query Match 30.5%; Score 239.4; DB 12; Length 828;
Best Local Similarity 63.3%; Pred. No. 1.9e-39;
Matches 455; Conservative 0; Mismatches 171; Indels 93; Gaps 2;
QY 85 cccacactgccccgcgaagcgccatctgtggtctctgctgctgctctctacacctg 144
Db 34 CCTCCACCCACCTGCAATGTGCGACTGAGGGTCTTCTGCTGCTGCTCCCTTGGCCT 93
QY 145 tgatccccctcgctgagctcctctatccatgatgagcccaagagagctccttggtc 204
Db 94 TCGTGTCT 153
QY 205 tcacagctccagagcctactccaagccttcagccgacttttctctgaaaggttaacctgc 264
Db 154 TCCGTGGGCTTCAGAGGCTTCTCCAAAGCTTTAGTCGACTGTTCTCTTAAATAATGACCTG 213
QY 265 ttcggggcatagacagcttattctctgcccccttcggacttcgggggctcctctgggaact 324
Db 214 TAGGAGACTTGGCAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 273
QY 325 accacaaagagagagacagcagcaccagctgggggaacacacacctctccagccacctcc 384
Db 274 TCCATCAGGAAGAGAACCCAGGAGCAGAAATGGCAACCATACCTCTCTCAGCCACCTAC 333
QY 385 agatcgac----- 393
Db 334 AGATAGACAAAGGTGACTGACAACACAGACAGGGAGGTGCACATCTCGGAGAAAGTCGAGG 393
QY 393 -----aaagtacccagagatgagaga 414
Db 394 CCTCCATTGAGCCAGAACCGGAACCCGGGAAGGGGACTGGAAGGTTCCCAAAAGTAGAGCAA 453
QY 415 agagggcctggtaccatccagaaaggccagcagcagcttcacacacacacacacacccccc 474
Db 454 AAGAGCCCCCGGTGCTGTGCAGAGAGGTTCACCGACAGCTTGCCACCCAGAGCCCGGCA-- 512
QY 475 ggggtgctcttggatcatatgaagctgcacggcgaggtccacacagagatgacctggagg 534
Db 512 -GTTGGCTTCTGGATCATGAAGATGCCAAGCGGAGGAGCCGCCGATGTCTCAGAGGATG 570
QY 535 ggggcaactggtcagcagagagcagacacccgctcagggccatccggatgagactccgca 594
Db 571 GAGGCGCTGGCTCATAGAAAAGCGACATCGCATCGAGGCCATCGGGATGGGCTCCGTG 630
QY 595 agggggaccacacagagcgtccctagaagaggggagccgagagctcctccacactccaggtgt 654
Db 631 GAGGCGCCCTGAGGAGACGCTTGGAGGATGGGTCCATATCCCCCAACACGCAAGCTGC 690
QY 655 ccccccgaagacccacttactgtatcactcctcagggcctctcggcagctgtagggtg 713
Db 691 CTGTCAAGAAAGACACACTTTTCTTCTACATCTCTCAGGCCATCCCCAACACAGCTGTAAAGTGGGG 749
RESULT 5
AC010524 82585 bp DNA HTG 26-JAN-2000
LOCUS Homo sapiens chromosome 19 clone LLNL-278H5, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
ACCESSION AC010524
VERSION AC010524.2 GI:6758800
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 82585)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 82585)
AUTHORS DOE Joint Genome Institute.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 26, 2000 this sequence version replaced gi:5882413.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1	851:	contig of 851 bp in length	*
852	1592:	gap of unknown length	*
1593	2551:	gap of unknown length	*
2552	3328:	contig of 777 bp in length	*
3329	3986:	gap of unknown length	*
3987	4885:	contig of 658 bp in length	*
4886	5646:	gap of unknown length	*
5647	6393:	contig of 747 bp in length	*
6394	6635:	contig of 242 bp in length	*
6636	6791:	contig of 156 bp in length	*
6792	7767:	contig of 976 bp in length	*
7768	8638:	contig of 871 bp in length	*
8639	9513:	contig of 875 bp in length	*
9514	10412:	contig of 899 bp in length	*
10413	11039:	contig of 627 bp in length	*
11040	11973:	contig of 934 bp in length	*
11974	12980:	contig of 1007 bp in length	*
12981	13903:	contig of 923 bp in length	*
13904	14856:	contig of 953 bp in length	*
14857	15392:	contig of 536 bp in length	*
15393	16293:	contig of 901 bp in length	*
16294	17173:	contig of 880 bp in length	*
17174	18116:	contig of 943 bp in length	*
18117	18739:	contig of 623 bp in length	*
18740	20791:	contig of 2052 bp in length	*
20792	21737:	gap of unknown length	*
21738	22690:	contig of 953 bp in length	*
22691	23645:	contig of 955 bp in length	*
23646	24245:	gap of unknown length	*
24246	24828:	contig of 583 bp in length	*
24829	25871:	contig of 1043 bp in length	*
25872	26570:	contig of 699 bp in length	*
26571	27470:	contig of 900 bp in length	*
27471	28388:	contig of 918 bp in length	*
28389	28552:	contig of 164 bp in length	*
28553	28626:	contig of 74 bp in length	*
28627	29553:	contig of 927 bp in length	*
29554	29810:	contig of 257 bp in length	*
29811	30718:	contig of 908 bp in length	*
30719	30803:	contig of 85 bp in length	*
30804	31489:	contig of 886 bp in length	*
31490	32509:	contig of 1020 bp in length	*
32510	33039:	contig of 530 bp in length	*
33040	33991:	contig of 952 bp in length	*
33992	34622:	contig of 631 bp in length	*
34623	36081:	contig of 1459 bp in length	*
36082	36258:	contig of 177 bp in length	*
36259	36500:	contig of 242 bp in length	*
36501	37155:	contig of 655 bp in length	*
37156	38197:	contig of 1042 bp in length	*
38198	39228:	contig of 1031 bp in length	*
39229	39434:	contig of 206 bp in length	*
39435	40418:	contig of 984 bp in length	*
40419	41807:	contig of 1389 bp in length	*
41808	43300:	contig of 1493 bp in length	*
43301	44327:	contig of 1027 bp in length	*
44328	44426:	contig of 99 bp in length	*
44427	45458:	contig of 1032 bp in length	*
45459	46774:	contig of 1316 bp in length	*
46775	47053:	contig of 279 bp in length	*
47054	47938:	contig of 885 bp in length	*
47939	49239:	contig of 1301 bp in length	*
49240	50184:	contig of 945 bp in length	*

AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT

DOE Joint Genome Institute.
Sequencing of Human Chromosome 19
Unpublished
2 (bases 1 to 82585)
DOE Joint Genome Institute.

Direct Submission
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Genome version replaced gl:5882413.

Center: Joint Genome Institute
Center Code: JGI
web site: <http://www.jgi.doe.gov>

 * NOTE: This record contains 87 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 *
 *

*	*	1	851:	contig of 851 bp in length
*	*			gap of unknown length
*	*	852	1592:	contig of 741 bp in length
*	*			gap of unknown length
*	*	1593	2551:	contig of 959 bp in length
*	*			gap of unknown length
*	*	2552	3328:	contig of 777 bp in length
*	*			gap of unknown length
*	*	3329	3986:	contig of 658 bp in length
*	*			gap of unknown length
*	*	3987	4885:	contig of 899 bp in length
*	*			gap of unknown length
*	*	4886	5646:	contig of 761 bp in length
*	*			gap of unknown length
*	*	5647	6393:	contig of 747 bp in length
*	*			gap of unknown length
*	*	6394	6635:	contig of 242 bp in length
*	*			gap of unknown length
*	*	6636	6791:	contig of 156 bp in length
*	*			gap of unknown length
*	*	6792	7767:	contig of 976 bp in length
*	*			gap of unknown length
*	*	7768	8638:	contig of 871 bp in length
*	*			gap of unknown length
*	*	8639	9513:	contig of 875 bp in length
*	*			gap of unknown length
*	*	9514	10412:	contig of 899 bp in length
*	*			gap of unknown length
*	*	10413	11039:	contig of 627 bp in length
*	*			gap of unknown length
*	*	11040	11973:	contig of 934 bp in length
*	*			gap of unknown length
*	*	11974	12980:	contig of 1007 bp in length
*	*			gap of unknown length
*	*	12981	13903:	contig of 923 bp in length
*	*			gap of unknown length
*	*	13904	14856:	contig of 953 bp in length
*	*			gap of unknown length
*	*	14857	15392:	contig of 536 bp in length
*	*			gap of unknown length
*	*	15393	16293:	contig of 901 bp in length
*	*			gap of unknown length
*	*	16294	17173:	contig of 880 bp in length
*	*			gap of unknown length
*	*	17174	18116:	contig of 943 bp in length
*	*			gap of unknown length
*	*	18117	18739:	contig of 623 bp in length
*	*			gap of unknown length

Query Match	22.6%	Score 177.8;	DB 45;	Length 82585;
Best Local Similarity	93.9%	pred. No. 2.9e-27;		
		0. mismatches	12;	
			Indels	0;
			Gaps	0;

80	Qy	agcctccccacctgcgcccccgaaggcggaatgctggtcgtgctgctgctctctctac	139
56874	Db	agcctccccacctgcgcccccgaaggcggaatgctggtgctgctgctctctac	66933
140	Qy	cctggtgatcccttcgcgtcgaactctatcatgatgctgacgcccaagagagctcctt	199
66934	Db	cctggtgatcccttcgcgtcgaactctatcatgatgctgacgcccaagagagctcctt	66993
200	Qy	gggtctcacaggcctccagagctactccaggtcttcagccgacttcttcgaaggtta	259
66994	Db	gggtctcacaggcctccagagctactccaggtcttcagccgacttcttcgaaggtta	67053
260	Qy	cctgcttcggggcatag	276
67054	Db	cgatgctgggggggatgg	67070

RESULT		6			
AC010524/c	DNA	82585 bp	H7G		
LOCUS					
DEFINITION	Homo sapiens chromosome 19 clone LNLR-278H5, LOW-PASS SEQUENCE SAMPLING.				
ACCESSION	AC010524				
VERSION	AC010524.2	GI:6758800			
KEYWORDS	HTG; HTGS_PHASE0.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 82585)				

* 18740 20791: contig of 2052 bp in length
* gap of unknown length
* 20792 21737: contig of 946 bp in length
* gap of unknown length
* 21738 22690: contig of 953 bp in length
* gap of unknown length
* 22691 23645: contig of 955 bp in length
* gap of unknown length
* 23646 24245: contig of 600 bp in length
* gap of unknown length
* 24246 24828: contig of 583 bp in length
* gap of unknown length
* 24829 25871: contig of 1043 bp in length
* gap of unknown length
* 25872 26570: contig of 699 bp in length
* gap of unknown length
* 26571 27470: contig of 900 bp in length
* gap of unknown length
* 27471 28388: contig of 918 bp in length
* gap of unknown length
* 28389 28552: contig of 164 bp in length
* gap of unknown length
* 28553 28626: contig of 74 bp in length
* gap of unknown length
* 28627 29553: contig of 927 bp in length
* gap of unknown length
* 29554 29810: contig of 257 bp in length
* gap of unknown length
* 29811 30718: contig of 908 bp in length
* gap of unknown length
* 30719 30803: contig of 85 bp in length
* gap of unknown length
* 30804 31489: contig of 686 bp in length
* gap of unknown length
* 31490 32509: contig of 1020 bp in length
* gap of unknown length
* 32510 33039: contig of 530 bp in length
* gap of unknown length
* 33040 33991: contig of 952 bp in length
* gap of unknown length
* 33992 34623: contig of 631 bp in length
* gap of unknown length
* 34623 36081: contig of 1459 bp in length
* gap of unknown length
* 36082 36258: contig of 177 bp in length
* gap of unknown length
* 36259 36500: contig of 242 bp in length
* gap of unknown length
* 36501 37155: contig of 655 bp in length
* gap of unknown length
* 37156 38197: contig of 1042 bp in length
* gap of unknown length
* 38198 39228: contig of 1031 bp in length
* gap of unknown length
* 39229 39434: contig of 206 bp in length
* gap of unknown length
* 39435 40418: contig of 984 bp in length
* gap of unknown length
* 40419 41807: contig of 1389 bp in length
* gap of unknown length
* 41808 43300: contig of 1493 bp in length
* gap of unknown length
* 43301 44327: contig of 1027 bp in length
* gap of unknown length
* 44328 44426: contig of 99 bp in length
* gap of unknown length
* 44427 45458: contig of 1032 bp in length
* gap of unknown length
* 45459 46774: contig of 1316 bp in length
* gap of unknown length
* 46775 47053: contig of 279 bp in length
* gap of unknown length
* 47054 47938: contig of 885 bp in length

* 47939 49239: contig of 1301 bp in length
* gap of unknown length
* 49240 50184: contig of 945 bp in length
* gap of unknown length
* 50185 51973: contig of 1789 bp in length
* gap of unknown length
* 51974 52854: contig of 881 bp in length
* gap of unknown length
* 52855 53768: contig of 914 bp in length
* gap of unknown length
* 53769 55173: contig of 1405 bp in length
* gap of unknown length
* 55174 56155: contig of 982 bp in length
* gap of unknown length
* 56156 58168: contig of 2013 bp in length
* gap of unknown length
* 58169 60209: contig of 2041 bp in length
* gap of unknown length
* 60210 61328: contig of 1119 bp in length
* gap of unknown length
* 61329 62639: contig of 1311 bp in length
* gap of unknown length
* 62640 63029: contig of 390 bp in length
* gap of unknown length
* 63030 63621: contig of 592 bp in length
* gap of unknown length
* 63622 65297: contig of 1676 bp in length
* gap of unknown length
* 65298 66599: contig of 1302 bp in length
* gap of unknown length
* 66600 66741: contig of 142 bp in length
* gap of unknown length
* 66742 67370: contig of 629 bp in length
* gap of unknown length
* 67371 68009: contig of 639 bp in length
* gap of unknown length
* 68010 68927: contig of 918 bp in length
* gap of unknown length
* 68928 70365: contig of 1438 bp in length
* gap of unknown length
* 70366 72855: contig of 2490 bp in length
* gap of unknown length

Query Match 10.3%; Score 81; DB 45; Length 82585;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cggatccggagtcctgacccggctgtgctagcataaagcgagccagaaga 60
|||||
Db 63132 CGAATCCGGAGTCCGGTGACCCGGCTGTGCTAGCATAAAGCGGAGAG 63073
Qy 61 agggcggggtatggagaag 81
|||||
Db 63072 AGGGCGGGGTATGGAGAG 63052

RESULT 7
SGKSACPG
LOCUS SGKSACPG 5364 bp DNA BCT 19-MAY-1994
DEFINITION S.griseus genes for ketosynthase, acyl carrier protein,
ketoreductase, cyclase and dehydrase.
ACCESSION X77865
VERSION X77865.1 GI:488770
KEYWORDS acyl carrier protein; cyclase; dehydrase; ketoreductase;
ketosynthase.
SOURCE Streptomyces griseus.
ORGANISM Streptomyces griseus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 5364)
AUTHORS Yu,T.W., Bibb,M.J., Revilli,W.P. and Hopwood,D.A.
TITLE Cloning, sequencing, and analysis of the griseusin polyketide


```
/db_xref="taxon:9913"
/tissue_type="calf thymus"
/clone_lib="lambda EMBL3A"
/clone="lambda F100"
2143..2148
2171..2305
2975..3114
/product="beta-lactoglobulin"
/notes="exon 1"
2975..3114
/notes="exon 2"
/product="beta-lactoglobulin"
3976..4049
/notes="exon 3"
/product="beta-lactoglobulin"
5167..5277
/notes="exon 4"
/product="beta-lactoglobulin"
5953..6057
/notes="exon 5"
/product="beta-lactoglobulin"
6279..6321
/notes="exon 6"
/product="beta-lactoglobulin"
6711..6893
/notes="exon 7"
/product="beta-lactoglobulin"
6869..6874
/notes="polyA site"
7372..7877
/notes="artiodactyl Alu-like repetitive sequence"
1551 a 2395 c 2356 g 1575 t
ORIGIN

Query Match 5.8%; Score 45.6; DB 3; Length 7877;
Best Local Similarity 48.5%; Pred. No. 3.9;
Matches 126; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 307 ggggccccctgggaactaccacaaagagagaccggagcaccagctggggacaaca 366
Db 6058 GTGAGCACCAGGCCACCCCTGCTCTGGGGCAGGAGCCACCCGCCAGGACCACT 6117
Qy 367 cccctccagccactccagatgacaaagtaccagagatggagagaagagcccttg 426
Db 6118 CCTCCCATGGTGACCCCGAGCTCCCGAGGCTCCCGGAGATGGAGCGGGTGACGG 6177
Qy 427 taccatccagaagggcagcagctccacacagaaactccatcccggtggccttct 486
Db 6178 CCCGAGGTGGCCCCCTCCCGAGGCTCCCGAGGCTCCCGAGGCTCCCGAGGCT 6237
Qy 487 ggatcattaaagctgccacggcgaggtccaccagagatgcctggagggcgccactggc 546
Db 6238 CCATCCTGACGCTCCCGCGGCGAGCTCCCTCCCGAGGAGTGGAGCGAGTGCACATCTAGG 6297
Qy 547 tcagcgagaagcagacgcg 566
Db 6298 TGAGCCCCCTGCGCGGCGCTC 6317

RESULT 9
BTBLGLOBB
LOCUS BTBLGLOBB 9432 bp DNA MAM 15-FEB-1995
DEFINITION B.taurus gene for beta-lactoglobulin variant B.
ACCESSION Z48305
VERSION Z48305.1 GI:669060
KEYWORDS beta-lactoglobulin.
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 9432)
AUTHORS Hyttinen,J.M., Korhonen,V.P., Myohanen,S. and Janne,J.
```

```

TITLE Bovine beta-lactoglobulin: cloning and expression in transgenic
        mice
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 9432)
AUTHORS Hyttinen,J.M.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-1995) Juha M. Hyttinen, A.I.Virtanen Institute,
        University of Kuopio, Kuopio, FIN-70211, Finland
FEATURES
        source
            1..9432
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone="clone BLC39"
                /tissue_type="sperm"
                /clone_lib="lambda GEM-11"
                2738..2742
                /notes="exon 1"
                2766..2900
                /notes="exon 1"
                2766..7494
                /notes="beta-lactoglobulin mRNA and introns"
                join(2811..2900,3572..3711,4575..4648,5768..5878,
                6554..6658,6881..6897)
                /codon_start=1
                /product="beta-lactoglobulin variant B precursor"
                /protein_id="CAA88303.1"
                /db_xref="GI:669061"
                /db_xref="SWISS-PROT:P02754"
                /translation="MKCLLALATCGAQLIVTQTMKGLDIQKVAGTWYSLAAMSD
                ISLNDQASAPLRVYVELKPTPEGDLFLLQKWENGCAQKKAIAETKIPAVFKIDA
                LNKVKVLVDYDKYKLLFCMNSAEPEQSLACQLVRTPEVDDEALEKFKALKALP
                MHIRLSFNPTOLEEQCHI"
                2811..2858
                /product="beta-lactoglobulin signal peptide"
                join(2859..2900,3572..3711,4575..4648,5768..5878,
                6554..6658,6881..6894)
                /product="beta-lactoglobulin variant B"
                3572..3711
                /notes="exon 2"
                4575..4648
                /notes="exon 3"
                5768..5878
                /notes="exon 4"
                6554..6658
                /notes="exon 5"
                6881..6922
                /notes="exon 6"
                7312..7494
                /notes="exon 7"
                7470..7475
                polyA_signal
                polyA_site
                7494
BASE COUNT 1898 a 2832 c 2743 g 1959 t
ORIGIN

Query Match 5.8%; Score 45.6; DB 3; Length 9432;
Best Local Similarity 48.5%; Pred. No. 3.7;
Matches 126; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 307 ggggccccctgggaactaccacaaagagagaccggagcaccagctggggacaaca 366
Db 6659 GTGAGCACCAGGCCACCCCTGCTCTGGGGCAGGAGCCACCCGCCAGGACCACT 6718
Qy 367 cccctccagccactccagatgacaaagtaccagagatggagagaagagcccttg 426
Db 6719 CCTCCCATGGTGACCCCGAGCTCCCGAGGCTCCCGGAGATGGAGCGGGTGACGG 6778
Qy 427 taccatccagaagggcagcagctccacacagaaactccatcccggtggccttct 486
Db 6779 CCCGAGGTGGCCCCCTCCCGAGGCTCCCGAGGCTCCCGAGGCTCCCGAGGCT 6838
Qy 487 ggatcattaaagctgccacggcgaggtccaccagagatgcctggagggcgccactggc 546
Db 6839 CCATCCTGACGCTCCCGCGGCGAGCTCCCTCCCGAGGAGTGGAGCGAGTGCACATCTAGG 6898
```

[illegible]

LOCUS	SC7A1	32039 bp	DNA	BCT	15-DEC-1998	
DEFINITION	Streptomyces coelicolor cosmid 7A1.					misc_feature
ACCESSION	AL034447					gene
VERSION	AL034447.1	GI:4007715				
KEYWORDS	50S ribosomal protein L28; 50S ribosomal protein L32; acy; acylphosphatase; acyltransferase; ammonium transporter; amt; ATP-dependent DNA helicase; chromosome associated protein; D-alanine-D-alanine ligase; ddla; formamidopyrimidine-DNA glycosylase; fpg; ftsY; glnB; glycerol-3-phosphate dehydrogenase; gpdA; KDO transferase; kdtB; methylase; nitrogen regulatory protein P-II; phosphomethylpyrimidine kinase; prokaryotic docking protein; recG; ribonuclease III; rnc; rpmB; rpmF; SRP54; sugar transporter; thiamine monophosphate kinase; thid; thil; transcriptional regulator.					CDS
SOURCE	Streptomyces coelicolor A3(2).					misc_feature
ORGANISM	Streptomyces coelicolor A3(2)					
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.					gene
AUTHORS	1 (bases 1 to 32039)					
JOURNAL	Murphy,L. and Harris,D.					CDS
REFERENCE	2 (bases 1 to 32039)					misc_feature
AUTHORS	Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.					
JOURNAL	Submitted (09-DEC-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge					gene
TITLE	CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK					
REFERENCE	3 (bases 1 to 32039)					misc_feature
AUTHORS	Redenbach,M., Kieser,H.M., Denapaitte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.					
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome					misc_feature
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)					
MEDLINE	97000351					misc_feature
COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC. Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/jun/cgi-bin/frameplot.pl . CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 7A1 lies between 1C2 and 2E1 on the AseI-B genomic restriction fragment. Location/Qualifiers 1. :32039 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226"					
FEATURES	source					misc_feature
Source	/clones="cosmid 7A1"					
	1. :112					misc_feature
	/note="Overlap with Streptomyces coelicolor cosmid 1C2 (EMBL:AL031124)." complement(97..735) /gene="SC7A1.01c" complement(97..735) /gene="SC7A1.01c"					
	/note="SC7A1.01c, unknown, len: 212aa; similar to TR:P95112 (EMBL:283018) hypothetical protein from Mycobacterium tuberculosis (214 aa) fasta scores; opt: 420, z-score: 422.3, E(): 3.1e-16, (43.5% identity in 207 aa overlap) and TR:Q33128 (EMBL:299263) hypothetical protein from Mycobacterium leprae (216 aa) fasta scores; opt: 386, z-score: 414.0, E(): 9.1e-16, (40.6% identity in 207 aa overlap)." codon_start=1 /transl_table=1 /product="hypothetical protein SC7A1.01c" /protein_id="CAA22400.1" /db_xref="GI:4007716"					misc_feature
	/translation="MQWTLVYVYKALAKAKSRSLSDTADDGLRPGCLALAFADQTVAAALACPAAADVAVTDDARAGRELALGAGVADDEPGGLNALAHAAVVRARPSVVALANADLPALRPAELARVLAATQFPRAFLPDAAGIGTLLTVAPGQELAFAGADSRARASGAVELRLDAVDSVRQDVTGDLRSALALGVGPRTAAVAARLLIAGQ" 879..1673 /gene="SC7A1.02" 879..1673 /gene="SC7A1.02"					
	/note="SC7A1.02, possible acyltransferase, len: 264aa; similar to several proposed acyltransferases eg. TR:O07809 (EMBL:297188) acyltransferase from Mycobacterium tuberculosis (251 aa) fasta scores; opt: 266, z-score: 384.5, E(): 4e-14, (29.7% identity in 229 aa overlap)." codon_start=1 /transl_table=1 /product="putative acyltransferase" /protein_id="CAA22401.1" /db_xref="GI:4007717"					misc_feature
	/translation="MPRRIGFWRLAAVICKPPLVVLIKRDRGMENIPAEGGFIYAVNNSHVDPFAYAYQYNTKVRPLAKSGFLKKGFAAMRGQTGPVIVRESYDALAFRAIDAIVERGECVAFYPTGTLTRDTPDPMPTAKTGAARVALQTGCPVTPVAGWGNELLPYAKPSVLPKTKHVLGAGPPVLSFYDREMTVLKTEATEVIMAAVTRQLEEIRGEKAPTVPDPRRIEQRRTQQAQSQVAPRTHGPQAEQST" 1670..2680 /gene="SC7A1.03" 1670..2680 /gene="SC7A1.03"					
	/note="SC7A1.03, gpdA, glycerol-3-phosphate dehydrogenase, len: 366aa; similar to many eg. SW:GPDA_BACSU glycerol-3-phosphate dehydrogenase from Bacillus subtilis (345 aa) fasta scores; opt: 874, z-score: 1221.4, E(): 0, (41.8% identity in 330 aa overlap). Contains a possible PS00017 ATP /GTP-binding site motif A (P-loop) and Pfam match to entry PF01210 NAD_Gly3P_dh, NAD-dependent glycerol-3-phosphate dehydrogenase , score 295.10, E-value 8.8e-85." codon_start=1 /transl_table=1 /product="glycerol-3-phosphate dehydrogenase" /protein_id="CAA22402.1" /db_xref="GI:4007718"					misc_feature
	/translation="MSKPVKAAVFGTSGWGTAFGTVLADACCEVTILNGRRALADAVNSTRTNPDYLPGEVPELNRATDAAEAARDADFTVLAVPSQTLRAGLADWTPLLAGPTVLVLSMKGVLSEGSAMRSEVIGDVAKVGAERIAVVTGPNLARETAARMPAAVACPD ETVARLQAACHTPYFRPNTDVGCELGAVKNVIGLAVGIADGMLGDNAGKSLITRGLAETTRLGVALGADPLTFSGLAGLGLVATCSSLSRNHTFTGLNGKMTLEETNAVTKQTAEGVKSCSVLDDLARRHGVDPITETVVAIVHEGKSPVAVKELMSRAKPERR" 1670..2647 /gene="SC7A1.03" /note="Pfam match to entry PF01210 NAD_Gly3P_dh, NAD-dependent glycerol-3-phosphate dehydrogenase , score					

misc_feature	295.10, E-value 8.8e-85"	Db	9522	GACGGGCTGCTACCGGCTTCGAGACCGGCTCCCTTCCCTTACCCCTCACCAGCGCCACGGC 9581
	2600..2623	QY	507	cggaggtccaccacagatgccctggaaggcgccactgctcagcgaagcgacaccgc 566
gene	/gene="SC7A1.03"	Db	9582	AAGGTCTCCGGGAGATCTTCGACGACCTCGCCACCCGACCCGATGCTGCTC 9641
	/note="PS00017 ATP /GTP-binding site motif A (P-loop)."	QY	567	ctcagagccatcgggatgactccgcaaggggaccaccaagagcgtcctagaagggg 626
CDS	2772..3941	Db	9642	CAGGCGAGTGGCGGCGGACAGACCTGCTGCGCCATGCTGCGCGTCTGTC 9701
	/gene="SC7A1.04"	QY	627	accgagagctctcccaactccaggtgtcccccgaagaaccacttactgtacatcctc 686
misc_feature	389aa; similar to many eg. SW:DDLA.SALTY	Db	9702	GACTCGGCGGAGGCGGTGATGCTGGCGCCACCGAGGTGCTGCCAGCAGCAC 9761
	D-alanine-D-alanine ligase from Salmonella typhimurium	QY	687	agggcctctcggcagctgtagggtgggaccgggagcaccctgcctg 734
gene	(363 aa) fasta scores; opt: 742, z-score: 1018.5, E(): 0,	Db	9762	CGSTCGTCTGCGAGATGCGGCGAGCTGCGCGGCGGCGATGCTG 9809
	(39.3% identity in 377 aa overlap). Contains PS00843	RESULT 14		
CDS	D-alanine--D-alanine ligase signature 1."	MMP4G		3768 bp DNA ROD
	/codon_start=1	LOCUS		Mouse MP4 gene for a proline-rich protein. 30-JAN-1992
misc_feature	/transl_table=11	DEFINITION		X58438
	/product="D-alanine-D-alanine ligase"	ACCESSION		MP4 gene; proline rich protein.
gene	/db_xref="GI:4007719"	VERSION		GI:53181
	/translation="MSTENLPQNPEQSPRRPRVAVFEGRSSEHGISVTVAGV	KEYWORDS		house mouse.
CDS	LAADRTRVDLPIGITDRGWTADPERMAITERTPDVEELAESTEGGVLLPVD	SOURCE		Mus musculus
	PANREVVSEPGSKALGEVDVFPVLHGYPYDGTQGLLELSGVPYVGVGLASA	ORGANISM		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
misc_feature	VQGDREYKAVFTSGLVGYPXVIRPREWEQDRSGARKKIVDFAGHGWLFPKPAR	REFERENCE		Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	AGSGITKTVDDLADLAEIARHDPKILVERALGREIECGVLEFEDCPRASVPA	AUTHORS		McDonald,C.J.
gene	EIPPESEHAYDFEAKYIDSTPGIVPAPLTAETAEVQRLAVAFDAASCGLVRAF	TITLE		Direct Submission
	FLTEDEFVINEINTMPGFTISMYPQMWQASGSYPVELVDRLVQALRRPTGLR"	JOURNAL		Submitted (15-MAR-1991) C.J. McDonald, Department of Molecular
CDS	3162..3197			Biology and Biotechnology, University of Sheffield, Western Bank,
	/gene="SC7A1.04"			Sheffield S10 2TN, UK
gene	/note="PS00843 D-alanine-D-alanine ligase signature 1."	REFERENCE		2 (bases 1 to 3768)
	complement(4119..4352)	AUTHORS		Roberts,S.G., Layfield,R., Bannister,A.J. and McDonald,C.J.
CDS	complement(4119..4352)	TITLE		Gene sequence of mouse B-type proline-rich protein MP4.
	/gene="SC7A1.05c"	JOURNAL		Transcriptional start point and an upstream phylogenetic footprint
misc_feature	77aa; similar to several putative /hypotheticals eg.			with ets-like and rel/NFkB-like elements
	TR:O28868 (EMBL:AE001007) proposed transcriptional			Eur. J. Biochem. 202 (3), 969-974 (1991)
gene	regulatory protein from Archaeoglobus fulgidus (77 aa)	JOURNAL		Location/Qualifiers
	fasta scores; opt: 123, z-score: 217.3, E(): 8.3e-05,	MEDLINE		1..3768
CDS	(31.9% identity in 69 aa overlap)."	FEATURES		/organism="Mus musculus"
	/codon_start=1	source		/strain="SV/129"
misc_feature	/transl_table=11			/db_xref="taxon:10090"
	/product="hypothetical protein SC7A1.05c"			/clone.lib="SV/129 in EMBL cos2"
gene	/protein_id="CAA22404.1"			/clone="3B"
	/db_xref="GI:4007720"	TATA_signal		628..633
CDS	/translation="MQAVILLIOTEYKASTVAETISKLPVGIQAEVDVTGVDIVRA	exon		654..745
	QSDTVDELGRWVAKVQVQVDGITRLTICPVHL"			/number=1
misc_feature	4586..5554	mRNA		join(654..745,1613..1648,2434..3262,3524..3593)
	4586..5554	gene		/gene="MP4"
CDS	4586..5554	prim_transcript		/gene="MP4"
	/gene="SC7A1.06"			/evidence=experimental
misc_feature	note="SC7A1.06, thiL, thiamine monophosphate kinase, len:			join(682..745,1613..1648,2434..3236)
	322aa; similar to many eg. TR:P95118 (EMBL:Z83018)	CDS		/codon_start=1
gene	hypothetical protein from Mycobacterium tuberculosis (333			/product="proline rich protein"
	aa) fasta scores; opt: 959, z-score: 1227.0, E(): 0,			/protein_id="CAA41344.1"
CDS	Query Match			/db_xref="GI:53182"
	5.4%; Score 42.4; DB 1; Length 32039;			/db_xref="MGI:961888"
misc_feature	Best Local Similarity			/translation="MLVILLTAALLVLSAAREDEETIYEDNSOLLEMGESQGYGH
	43.2%; Pred. No. 13;			HFKPPGNGPMPRPSPSDNDDDDSEEDVNRGPPHPPHSGNHHPQQGDAH
CDS	Matches 202; Conservative 0; Mismatches 266; Indels 0; Gaps 0;			GPPRPGNQGPSPGPPPGSSQSRPPQPGNQGPQPGPPQPGPPQPGPPQPGPP
				GPQRPFPQPGNQGPQPGPPPPPPGPNQGGPPQPGPPQPGPPQPGPPQPGPP
misc_feature	QY 267 cggggcatagacagtattcttgcggccatggacttcggcgccctccctcgggaactac 326			
	Db 9342 CGGGGCTGCTCCCTGCGCGAGCCCTGCTCAAGATCTACCGCCGACACCAAGGCC 9401			
misc_feature	QY 327 cacaagagagaacagagaccagctggggaacacacccctctccagccactccag 386			
	Db 9402 GACATCGAGGACCGCGCCGCTGAAGTGGGACGAGGCTTCGTCCTCCAGTGC 9461			
misc_feature	QY 387 atcgacaaggtaccagagtgaggaagaggccctgggtaccatccagaaggaacag 446			
	Db 9462 CTCGCCCGCGCGCACCGAGTCCCAGCTCCCGCGCTCCCGAGGAAGCCGCGCC 9521			
misc_feature	QY 447 gacagcttccacagaaactccatcccggtggcctctctgattatgaagctgcacgg 506			
	Db 447 gacagcttccacagaaactccatcccggtggcctctctgattatgaagctgcacgg 506			

QAPPRGNGQQGPPPGQGGPPRTGNQGGPPPGQGGPPRFGNQGGPPPPQGGPQRRPPQ
PGNHQGPQGHGNNQPSYLSIFA"

intron 746..1612
/gene="MP4"
/number=1
exon 1613..1648
/gene="MP4"
/number=2
intron 1649..2433
/gene="MP4"
/number=2
exon 2434..3262
/gene="MP4"
/number=3
intron 3263..3523
/gene="MP4"
/number=3
exon 3524..3593
/gene="MP4"
/number=4
polya_signal 3576..3581
/gene="MP4"

BASE COUNT 1186 a 872 c 704 g 1006 t
ORIGIN

Query Match 5.4%; Score 42; DB 12; Length 3768;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 84; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 539 ccaactggctcagcagaagcgacaccgcctgcaggccatcccggtatggactccgcaagg 598
|||||
Db 3027 CCACAGGGCCCTCAGGACCTCTAGGACTTGGAAACCAAGGCCACCCACACAGGGA 3086
QY 599 gaccacagagcgtctctagagaggggaccgagagctctccatccagcgtgtcccc 658
|||||
Db 3087 GGCCTTCAGGAGCCTCTAGGCTTGGAAACCAAGGCCACCCACACAGGAGGCCCT 3146
QY 659 ccgaagacccacttactgtacatctctcagggcc 692
|||||
Db 3147 CAGAGACCTCTCAGCCAGCAACCAATCAGGCC 3180

RESULT 15
AB018254 6706 bp mRNA PRI 16-JUN-1999
LOCUS
DEFINITION Homo sapiens mRNA for KIAA0711 protein, complete cds.
ACCESSION AB018254
VERSION AB018254.1 GI:3882142
KEYWORDS
SOURCE Homo sapiens adult male brain cdna to mRNA, clone_lib:pBluescriptII
SK plus clone:hg00358.
ORGANISM Homo sapiens

REFERENCE 1 (sites)
AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
JOURNAL DNA Res. 5 (5), 277-286 (1998)
MEDLINE 99087487

REFERENCE 2 (bases 1 to 6706)
AUTHORS Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) to the DDBJ/EMBL/GenBank databases. Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 252-0812, Japan
(E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES
source 1..6706
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hg00358"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
964..2835
/gene="KIAA0711"
964..2835
/gene="KIAA0711"
/codon_start=1
/product="KIAA0711 protein"
/protein_id="BAA34431.1"
/db_xref="GI:3882143"

gene
CDS

/translation="MEHAVPCVLYPGTEPGAEGSESGAASPAQTPCSLGSALCFSS
SGEESPQSLASAAEAGATSPSSGGPRVVERQWAGSAGAAPELASPEERACPEE
PAAPSPPEPRVLEWEDPASPEEPGEPAFPFGFAGVYGEDVLVLEVSRRRLRAHKAVLAA
RSDYFRASRDVLRVQGVSLFALRLLADAYSGRMAGVYRPNDAVAVAGARLQLP
AAQRATDVGQPSLANCYLVAALGPGAGERAGSRPQSPGSDADARGDAVYCFHAAA
SGAERDILLRRRLRAGRAHLLAAALGPGAGERAGSRPQSPGSDADARGDAVYCFHAAA
GEWRELTRLPAGAPARGCGCLVLYNLFVAGGVAPAGPDGRARPSDQVFCYNPATDSW
SAVRPLQRARSQRLALLDGHLYAVGECULLSVERTDPRADRWAPVAPLPRGAFVAH
EATTCHEIYVGGSLFYRLKYDPRRDEWECPCSSRSRSDMVALDGFYTFDL
GSRGEAQAAGPSGVSYRYHLAKQWPCVAPLRLPGPTGLQPFRCALDGAITYCVS
RAGTWRFQPARREGAGDAGGGGFEALGPLDVRGLVLPFALSLEKPPRGGAP"

BASE COUNT 1478 a 1736 c 1957 g 1535 t
ORIGIN

Query Match 5.4%; Score 42; DB 9; Length 6706;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 105; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 440 ggcacggcagagcttccacacagaaatccatccccgggtggcttcttggaatcattagct 499
|||||
Db 2124 GGCACGGACAGCTGGAGCGCGCTGAGGCCCTCGCCAGCGCGCTCGCAGCTGCGGCT 2183
QY 500 gccacggcggaggtccaccagcagatccctggagggcgccactggctcagcagaagcg 559
|||||
Db 2184 GCTGGCCCTGGACGGTCACTCTACGCCGTGGGGGGGAGTGCTGTCTCAGCGTGGAGCG 2243
QY 560 acaccgctgcaggccatccggatgactccgcgaaggggaccacacaaagcagctcctaga 619
|||||
Db 2244 CTAGACCCCGCGCGACCGCTGGGCCCGCTGGCGCCGCTGCCCGGGCGGCTTCGC 2303
QY 620 agaggggaccgagagctctccactccag 649
|||||
Db 2304 CGTGGCGCATGAGGCCACCACTGCCACGG 2333

Search completed: May 24, 2000, 00:42:13
Job time: 13284 sec

us-09-215-435-1118.rge

Wed May 24 09:19:45 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:02:10 ; Search time 282.38 seconds
(without alignments)
695.519 Million cell updates/sec

Title: US-09-215-435-118

Perfect score: 785

Sequence: 1 cgggaatccggagtcggtg.....tatggaataaagtctttc 785

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	660	84.1	928	1 V07910	Human cysteine-rich
2	384.6	49.0	464	1 X51459	Human secreted protein
3	41	5.2	1137	1 Q33061	Plasmodium vivax c
4	40.4	5.1	459	1 V34179	Human secreted pro
5	39.4	5.0	1532	1 V45173	Synthetic HIV gag
6	39	5.0	1146	1 Q83793	S. avermitilis BCK
7	39	5.0	2728	1 Q83797	S. avermitilis bkd
8	39	5.0	15672	1 Q10613	Rianodin receptor
9	38.8	4.9	2538	1 X21398	Central fragment o
10	38.8	4.9	7063	1 X24751	Human interleukin-
11	38.4	4.9	2357	1 Q62177	Human TLE-3 gene.
12	38.4	4.9	2357	1 Q45335	Human TLE-3 gene.
13	37.8	4.8	470	1 V80252	Modified Brazil nu
14	37.8	4.8	474	1 V80254	Modified Brazil nu
15	37.8	4.8	1908	1 N71064	Gene encoding Plas
16	37.4	4.8	114955	1 X53491	Human adenosine A1
17	37.2	4.7	384	1 X41095	Human secreted pro
18	37.2	4.7	2499	1 N90366	Synthetic human 5-
19	37	4.7	2417	1 Q56106	DNA encoding a lip
20	36.8	4.7	1142	1 X27314	Human secreted pro
21	36.8	4.7	2512	1 Q92369	TfII DNA-polymeras
22	36.6	4.7	2058	1 V07560	Neocallimastix pat
23	36.6	4.7	3177	1 T16474	B virus gB glycopr
24	36.6	4.7	3177	1 V33167	Simian herpesvirus
25	36.6	4.7	15079	1 Q91580	S. clavuligerus cl
26	36.4	4.6	2092	1 T35762	Human non-conventi
27	36	4.6	2238	1 T06476	Bovine calcium-act
28	36	4.6	2943	1 T16480	SAB virus gB glyco
29	35.8	4.6	1419	1 T10928	Nucleotide sequenc
30	35.8	4.6	1419	1 V62709	Maize male fertill
31	35.8	4.6	1419	1 V73926	Maize MS45 cDNA. N
32	35.8	4.6	1419	1 X00482	Zea mays male fert
33	35.8	4.6	2165	1 T59702	Endotoxin Cryv gen
34	35.8	4.6	2744	1 Q98470	MISPI-containing p

ALIGNMENTS

RESULT 1

V07910
ID V07910 standard; cDNA; 928 BP.
AC V07910;
DE Human cysteine-rich secreted protein-like-N cDNA.
KW CRSP-like-N; cysteine-rich secreted protein; tumour; cancer;
KW signal transduction; cell differentiation; cell proliferation;
KW human; ss.
OS Homo sapiens.
PH Key Location/Qualifiers
FT CDS 75..803 /*tag= a
FT WO9846755-A1.
PD 22-OCT-1998.
PF 16-APR-1998; U07894.
PR 20-JAN-1998; US-009802.
PR 16-APR-1997; US-843704.
PR 17-APR-1997; US-842898.
PR 15-JAN-1998; US-071589.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI McCarthy SA;
DR WPI: 98-568730/48.
DR P-PSDB; W73020.
PT New isolated cysteine-rich secreted proteins - used to develop
PT products for treating, e.g. hyperproliferative disorders, cancers,
PT wounds, infectious lesions, degenerative lesions or demyelating
PT diseases
PS Example 2; Page 107-108; 142pp; English.
CC This full-length cDNA clone codes for a novel human protein (see
CC W73020), designated CRSP-like-N, that shows homology to a novel,
CC claimed human cysteine-rich secreted protein 1 (CRSP-1, see
CC W73016). CRSP-like-N was identified using the nucleotide
CC sequence encoding the N-terminal unique region of CRSP-1 to
CC search a protein sequence database. Clone AA397836 was identified
CC and was subsequently isolated from the IMAGE collection and
CC sequenced fully to define the entire hCRSP-like-n sequence.
CC CRSP-1, -2, -3 and -4 nucleic acid sequences (see V07906-09) and
CC polypeptides (see W73016-19) are claimed. These can be used in
CC diagnostic, screening and therapeutic methods of the invention e.g.
CC for treating hyperproliferative disorders, cancers, wounds, and in
CC infectious or degenerative lesions and demyelating diseases, and in
CC drug screening.
SQ Sequence 928 BP; 229 A; 291 C; 257 G; 151 T;

Query Match 84.1%; Score 660; DB 1; Length 928;
Best Local Similarity 89.1%; Pred. No. 1.2e-143;
Matches 763; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
Qy 23 ccggcgtgtctagcataaagcgagcccaagagggcggtatggagagc 82
Db 26 CCGGCTGTGTCTAGCATAAAGCGGAGCCCAAGAGGGGGGTATGGAGAGC 85
Qy 83 ctccccacctgcccccgcaagcggcacatctgtgctgtgtgtctctctctaccc 142
Db 86 CTCCCCACTGCCCGCCGCAAGCGGCATCTGCTGCTGCTGCTCTCTACCT 145

cDNA encoding John
Sequence encoding
Total DNA sequence
S. aureofaciens DN
Sequence encoding
New DNA sequence 1
New DNA sequence 1
Modified Brazil nu
Modified Brazil nu
Modified Brazil nu
Modified Brazil nu

QY	143	ggtgacccctccgctgcagctcctatccatgatctgacgcccgaagagagctccttggg	202
DB	146	ggtgacccctccgctgcagctcctatccatgatctgacgcccgaagagagctccttggg	205
QY	203	tctcaaggctccagagctcctcaaggcttcacgagcttctcctgaaggttaacct	262
DB	206	tctcaaggctccagagctcctcaaggcttccagagcttctcctgaaggttaacct	265
QY	263	gcttcggggcatagacagcttattctctgcccccatggacttcctggggcctccctgggaa	322
DB	266	gcttcggggcatagacagcttattctctgcccccatggacttcctggggcctccctgggaa	325
QY	323	ctaccacaaagagagagaccagagaccagctgagggaacacacacctctccagcaacct	382
DB	326	ctaccacaaagagagagaccagagaccagctgagggaacacacacctctccagcaacct	385
QY	383	ccagatcgac-----	393
DB	386	ccagatcgac-----	445
QY	393	-----aagttaccagagtgga	409
DB	446	ggcatccattcaaccagggaggagcttcgaggggtgatttgaaggtaccagagtgga	505
QY	410	ggagaaggagccctgtgtaccatccagaagggccagagcttccacagagaactcca	469
DB	506	ggagaaggagccctgtgtgtaccatccagaagggccagagcttccacagagaactcca	565
QY	470	tcccggtggtcctcttgatcattatgaagctgcccagcggaggttccaccagagatgcct	529
DB	566	tcccggtggtcctcttgatcattatgaagctgcccagcggaggttccaccagagatgcct	625
QY	530	ggaggggccactggctcagcagagagcagcagctgcagggcactccggagtgact	589
DB	626	ggaggggccactggctcagcagagagcagcagctgcagggcactccggagtgact	685
QY	590	ccgcaaggagggccacacagagcgtctagagaggggacagagagctcctcccaactccag	649
DB	686	ccgcaaggagggccacacagagcgtctagagaggggacagagagctcctcccaactccag	745
QY	650	gctgtcccccggaaagaccacttactgtatcactcagggccctcctcggagctgtggg	709
DB	746	gctgtcccccggaaagaccacttactgtatcactcagggccctcctcggagctgtggg	805
QY	710	gtgggagcggggagacacactcctgtaccccatcagacccctgcccaagcaccatag	769
DB	806	gtgggagcggggagacacactcctgtaccccatcagacccctgcccaagcaccatag	865
QY	770	gaataaaagtcttttc 785	
DB	866	gaataaaagtcttttc 881	
RESULT	2		
ID	X51459		
AC	X51459		
DT	21-JUN-1999	(first entry)	
DE	Human	secreted protein 5' EST SEQ ID NO:38	
KW	Human	secreted protein; EST; expressed sequence tag; diagnosis;	
KW	forensic	gene therapy; chromosome mapping; signal peptide;	
KW	upstream	regulatory sequence; cytokine activity; cell proliferation;	
KW	differentiation	haematopoiesis regulation; tissue growth regulation;	
KW	reproductive	hormone regulation; chemotactic; chemokinetic; haemostatic;	
KW	thrombolytic	anti-inflammatory; tumour inhibition; ds.	
OS	Homo sapiens.		
PN	WO906549-A2		
PD	11-FEB-1999		
PF	31-JUL-1998	IB1231.	
PR	01-AUG-1997	US-905279.	
PA	(GENSET)	GENSET.	
PI	Duclet A, Dumas Milne Edwards J, Lacroix B;		
DR	P-PSDB: V12681.		
PT	New nucleic acids encoding human secreted proteins - obtained from		
PS	CDNA libraries derived from testis, ovary, uterus and spleen tissue		
CC	Claim 1; Page 171-172; 522pp; English.		
CC	X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human		
CC	secreted proteins, and encode the proteins given in Y12681 to Y12913,		
CC	respectively. The proteins given represent the signal peptide and an		
CC	N-terminal fragment of a secreted protein. The nucleic acid sequences		
CC	can be used for producing secreted human gene products. They can also		
CC	be used to develop products for diagnosis and therapy. The proteins		
CC	obtained may have cytokine activity, cell proliferation/differentiation		
CC	activity, haematopoiesis regulating activity, tissue growth regulating		
CC	activity, reproductive hormone regulating activity, chemotactic/		
CC	chemokinetic activity, haemostatic and thrombolytic activity, receptor/		
CC	ligand activity, anti-inflammatory activity, tumour inhibition activity		
CC	or other activities. The products can be used in forensic, gene therapy		
CC	and chromosome mapping procedures. The sequences can also be used for		
CC	obtaining corresponding promoter sequences. The nucleic acids encoding		
CC	the signal peptide can be used for directing extracellular secretion of		
CC	a polypeptide or the insertion of a polypeptide into a membrane, or		
CC	importing a polypeptide into a cell.		
SQ	Sequence 464 BP; 102 A; 146 C; 132 G; 81 T;		
Query Match	49.0%;	Score 384.6;	DB 1; Length 464;
Best Local Similarity	94.4%;	Pred. NO. 3e-80;	
Matches 407;	Conservative 1;	Mismatches 22;	Indels 1; Gaps 1;
QY	1	cggaaatccggaggtccggtgacccggctgtgtctagcataaaaggcggagccagaaga	60
DB	20	cggaaatccggaggtccggtgacccggctgtgtctagcataaaaggcggagccagaaga	78
QY	61	agggggcggtgtatgggagagagcctcccccacgtgccccgcaaggcggtatctgctggtcc	120
DB	79	agggggcggtgtatgggagagagcctcccccacgtgccccgcaaggcggtatctgctggtcc	138
QY	121	tgctgtgctctctctacccctgggtgagtcctccctccgctcagctcctatccatgatcgtg	180
DB	139	tgctgtgctctctctacccctgggtgagtcctccctccgctcagctcctatccatgatcgtg	198
QY	181	acgcccagaagagctccttgggtctcacagggctccacagagctcctcctcctcctcagcc	240
DB	199	acgcccagaagagctccttgggtctcacagggctccacagagctcctcctcctcctcagcc	258
QY	241	gactttctctgaaggttaacctgtcttcgggggatacagagcttattctctgcccccatgg	300
DB	259	gactttctctgaaggttaacctgtcttcgggggatacagagcttattctctgcccccatgg	318
QY	301	acttcggggcctccctgggaactaccacaaagagagagagagagagagagagagagag	360
DB	319	acttcggggcctccctgggaactaccacaaagagagagagagagagagagagagagag	378
QY	361	acaacacccctctccagccacctccagatcgacagaggtaccagaggtggagagagagag	420
DB	379	acaacacccctctccagccacctccagatcgacagaggtaccagaggtggagagagagag	438
QY	421	cccttggtaccc 431	
DB	439	tgctgatctcc 449	
RESULT	3		
ID	Q33061		
AC	Q33061		
DT	06-MAY-1993	(first entry)	
DE	Plasmodium vivax circumsporozoite (CS) gene.		
KW	Malaria; vaccine; immunogenic polypeptide; ss.		
OS	Plasmodium vivax.		
PF	Key	Location/Qualifiers	
FT	cds	1..1137	
FT		/*tag= a	

```

FT misc_feature /note= "circumsporozoite protein"
FT 247. .945
FT /*tag= b
FT /note= "DNA fragment encoding immunogenic polypeptide"
FT 292. .318
FT repeat_unit /*tag= c
FT /note= "encodes immunogenic repetitive sequence"
FT 319. .345
FT /*tag= d
FT /note= "encodes immunogenic repetitive sequence"
FT 346. .372
FT /*tag= e
FT /note= "encodes immunogenic repetitive sequence"
FT 373. .399
FT /*tag= f
FT /note= "encodes immunogenic repetitive sequence"
FT 400. .426
FT /*tag= g
FT /note= "encodes immunogenic repetitive sequence"
FT 427. .453
FT /*tag= h
FT /note= "encodes immunogenic repetitive sequence"
FT 454. .480
FT /*tag= i
FT /note= "encodes immunogenic repetitive sequence"
FT 481. .507
FT /*tag= j
FT /note= "encodes immunogenic repetitive sequence"
FT 508. .534
FT /*tag= k
FT /note= "encodes immunogenic repetitive sequence"
FT 535. .561
FT /*tag= l
FT /note= "encodes immunogenic repetitive sequence"
FT 562. .588
FT /*tag= m
FT /note= "encodes immunogenic repetitive sequence"
FT 589. .615
FT /*tag= n
FT /note= "encodes immunogenic repetitive sequence"
FT 616. .642
FT /*tag= o
FT /note= "encodes immunogenic repetitive sequence"
FT 643. .669
FT /*tag= p
FT /note= "encodes immunogenic repetitive sequence"
FT 670. .696
FT /*tag= q
FT /note= "encodes immunogenic repetitive sequence"
FT 697. .723
FT /*tag= r
FT /note= "encodes immunogenic repetitive sequence"
FT 724. .750
FT /*tag= s
FT /note= "encodes immunogenic repetitive sequence"
FT 751. .777
FT /*tag= t
FT /note= "encodes immunogenic repetitive sequence"
FT 778. .804
FT /*tag= u
FT /note= "encodes immunogenic repetitive sequence"
FT 804. .830
FT /*tag= v
FT /note= "contains 19 tandem repeats"
FT 830. .856
FT US5171843-A.
FT 15-DEC-1992.
FT 30-MAR-1988; US-649903.
FT 12-SEP-1984; US-754645.
FT 09-JUL-1985; US-115634.
FT 26-OCT-1987; US-175112.
FT 30-MAR-1988; US-175112.
FT (UYN ) UNIV NEW YORK STATE.
FT Nussenzweig V;

DR WPI; 93-008637/01.
DR P-PSDB; R30609.
PT New recombinant produced immunogenic poly:peptide - incorporating
PT part of P. vivax circumsporozoite protein and another part
PT conserved among different species
PS Disclosure; Page 14; 19pp; English.
CC The sequence is that of the plasmodium vivax circumsporozoite (CS)
CC gene, the CS protein encoded by the gene contains a repeat
CC immunodominant epitope comprising a central region of 19 tandem
CC repeats of the sequence DRA(D/A)QPPAG. This immunogenic polypeptide
CC is immunochemically reactive with monoclonal antibodies against P.
CC vivax CS protein and is useful in a vaccine against malaria. It may
CC be prepared by recombinant DNA techniques in yeast.
SQ Sequence 1137 BP; 393 A; 222 C; 348 G; 174 T;

Query Match 5.2%; Score 41; DB 1; Length 1137;
Best Local Similarity 47.5%; Pred. No. 0.5;
Matches 122; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 321 aactacacaaaggaggaaccaggagcaccagctgggggaacacacacccctctccagccac 380
DB 362 AACAGCAGGAGATAGACAGCTGGACAAACAGCAGGAGATAGACAGATGAGACAGCCAG 421
QY 381 ctccagatcgcgaaggtaccagcagatggaggaaggagccctgggtaccatccagaag 440
DB 422 CAGGAGACAGCAGCAGATGGACGCCAGCAGGAGACAGAGATGGACACACAGCAGGAG 481
QY 441 gccacggacagcttccacacagaactccatccccgggtggcctctctggtatcataagctg 500
DB 482 ACAGAGCAGATGGACAAACAGCAGCGTGTATAGACAGCTGGACAAACAGCAGGATAGAG 541
QY 501 ccagggcggaggtcccaccagcagatgcctggaggcggccactggctcagcagagaagcga 560
DB 542 CAGCTGGACAAACAGCAGGAGATAGACAGATGAGACGCCAGCAGGAGATAGAGCAGCTG 601
QY 561 caccgcctgcaggccat 577
DB 602 GACAGCCAGCAGGAGAT 618

RESULT 4
ID V34179 standard; DNA; 459 BP.
AC V34179; 1999 (first entry)
DE Human secreted protein gene 26 clone HTLEV12.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN W09839446-A2.
PD 11-SEP-1998.
PF 06-MAR-1998; U04492.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043315.
PR 11-APR-1997; US-043568.

```


CC involving: (1) identifying placement of codons for proper reading
 CC frame in the gag gene; (2) comparing a wild-type codon for observed
 CC frequency of use by human genes; (3) if necessary, replacing the
 CC codon with an optimal codon for high expression in human cells; (4)
 CC repeating (1)-(3) until the entire gene segment has been replaced;
 CC (5) checking the new gene sequence for undesired sequences (e.g.
 CC ATTTA sequences) and eliminating these; and (6) assembling
 CC synthetic gene segments and testing for improved expression.
 CC The synthetic gene can be used as a polynucleotide vaccine which
 CC provides effective immunoprophylaxis against HIV infection through
 CC neutralising antibody and cell-mediated immunity. The invention
 CC provides nucleotides which, when directly introduced into a
 CC vertebrate in vivo, including mammals such as primates and humans,
 CC induce the expression of encoded proteins within the animal.
 SQ Sequence 1532 BP; 357 A; 446 C; 476 G; 253 T;

Query Match 5.0%; Score 39.4; DB 1; Length 1532;
 Best Local Similarity 56.6%; Pred. No. 1.2;
 Matches 73; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 289 ctgcccccatggaattccggggcctcctgggaactaccacaagagagaaaccaggagc 348
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 368 CTGCACAGGCAACTCCAGCCAGGTGTCACAGAACTACCCCATTTGCGAGAACTCCAGG 427
 Qy 349 accagctggggaacaacacctctccagccacctccagatcgacaaggtaccagagtg 408
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 428 GCCAGATGTCACAGGCCATCTCCCGGACCCCTGATGCTGGTGAAGGTGGTGG 487
 Qy 409 aggaagaagg 417
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 488 AGGAGAAGG 496

RESULT 6

ID Q83793 standard; cDNA; 1146 BP.
 AC Q83793;
 DT 05-SEP-1995 (first entry)
 DE S. avermitilis BCKDH El-alpha subunit.
 KW Branched-chain alpha-keto-acid-dehydrogenase; BCKDH; bkd gene;
 KW avermectin; antibiotic; acaricide; anthelmintic; insecticide;
 KW nematocide; pesticide; ds.
 OS Streptomyces avermitilis.
 FH Key Location/Qualifiers
 FT cds 1..1146
 FT /*tag= a
 PN WO9504150-A.
 PD 09-FEB-1995.
 PF 30-MAY-1994; IB0127.
 PR 30-JUL-1993; US-100518.
 PA (PFIZ) PFIZER INC.
 PI Denoya CD;
 DR WPI; 95-082233/11.
 DR P-PSDB; R69623.
 PT Streptomyces branched-chain alpha-keto-acid dehydrogenase DNA -
 PT used to improve prodn. of natural avermectin cpds and to produce
 PT novel avermectin cpds
 PS Claim 5; Page 42; 65pp; English.
 CC Novel branched-chain alpha-keto-acid-dehydrogenase (BCKDH) genes
 CC (bkd) from Streptomyces avermitilis were cloned using PCR and
 CC homology probing. DNAs encoding the BCKDH El-alpha, El-beta, and
 CC N-terminal and internal portions of the E2 subunit (Q83793-96)
 CC were obtained. A genomic sequence including the El-alpha, El-beta
 CC and E2 (partial) bkd ORFs was also isolated (Q83797). Manipulation
 CC of bkd genes allows the enhanced production of natural or novel
 CC avermectins.
 SQ Sequence 1146 BP; 163 A; 465 C; 392 G; 126 T;

Query Match 5.0%; Score 39; DB 1; Length 1146;
 Best Local Similarity 49.7%; Pred. No. 1.4;
 Matches 99; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 250 tgaaggtaacctgtctcggggcataagacagtatttctgtcccccatgaacttcggg 309
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 917 TGACCGAAGCGGGCTGCTGACGAGGACGCATCGGGGCGCCCGGAGGACCGCGAGG 976
 Qy 310 gcctccctgggaactaccacaaaagagagaacaggagcaccagctggggaacaacacc 369
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 977 CGATGGCGCGGACCTGCGCGCACGCATGAACAGGATCCGCGCTGGACCCCATGGACC 1036
 Qy 370 tctccagccactccagatcgacaaggtaccagagtgagaggaagagagccctgtgac 429
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1037 TGTTCGCCCATGTGTATGCCGAGCCACCCCGCAGCTCGGGAGCAGGAAGCCAGTTGC 1096
 Qy 430 ccattccagaagccacgga 448
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1097 GGGCGAGCTGGCAGCGGA 1115

RESULT 7

ID Q83797 standard; cDNA; 2728 BP.
 AC Q83797;
 DT 05-SEP-1995 (first entry)
 DE S. avermitilis bkd gene region.
 KW Branched-chain alpha-keto-acid-dehydrogenase; BCKDH; bkd gene;
 KW avermectin; antibiotic; acaricide; anthelmintic; insecticide;
 KW nematocide; pesticide; ds.
 OS Streptomyces avermitilis.
 FH Key Location/Qualifiers
 FT rbs 390..395
 FT /*tag= a
 FT cds 403..1548
 FT /*tag= b
 FT /product= El-alpha subunit
 FT rbs 1607..1614
 FT /*tag= c
 FT cds 1622..2626
 FT /*tag= d
 FT /product= El-beta subunit
 FT rbs 2610..2615
 FT /*tag= e
 FT cds 2626..2728
 FT /*tag= f
 FT /product= E2 N-terminal sequence
 PN WO9504150-A.
 PD 09-FEB-1995.
 PF 30-MAY-1994; IB0127.
 PR 30-JUL-1993; US-100518.
 PA (PFIZ) PFIZER INC.
 PI Denoya CD;
 DR WPI; 95-082233/11.
 DR P-PSDB; R69623; R69624; R69625.
 PT Streptomyces branched-chain alpha-keto-acid dehydrogenase DNA -
 PT used to improve prodn. of natural avermectin cpds and to produce
 PT novel avermectin cpds
 PS Claim 5; Page 44-45; 65pp; English.
 CC Novel branched-chain alpha-keto-acid-dehydrogenase (BCKDH) genes
 CC (bkd) from Streptomyces avermitilis were cloned using PCR and
 CC homology probing. DNAs encoding the BCKDH El-alpha, El-beta, and
 CC N-terminal and internal portions of the E2 subunit (Q83793-96)
 CC were obtained. A genomic sequence including the El-alpha, El-beta
 CC and E2 (partial) bkd ORFs was also isolated (Q83797). Manipulation
 CC of bkd genes allows the enhanced production of natural or novel
 CC avermectins.
 SQ Sequence 2728 BP; 388 A; 1005 C; 963 G; 372 T;

Query Match 5.0%; Score 39; DB 1; Length 2728;
 Best Local Similarity 49.7%; Pred. No. 1.7;
 Matches 99; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 250 tgaaggtaacctgtctcggggcataagacagtatttctgtcccccatgaacttcggg 309
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

[illegible]

566 cctgcagcccatccgggagtgatccactccgaag 596
||||| ||||| ||||| |||||
6961 CCTGCACGGCAGTACGATGGCTCGGCGAG 6991

RESULT 9

X21398/c
ID X21398 standard; DNA; 2538 BP.
X21398;
21-MAY-1999 (first entry)
DE Central fragment of AbsA operon from S.coelicolor.
DE AbsA; operon; open reading frame; histidine kinase; response regulator;
KW mutation; insertion; antibiotic; actinorhodin; undecylprodigiosin; ss.
KW Streptomyces coelicolor.
OS Streptomyces coelicolor.
FH Key Location/Qualifiers
FT CDS 136..1851
FT FT /*tag= a
FT FT /gene= "AbsA1"
FT FT /product= "histidine kinase"
FT FT 1870..2538
FT FT /*tag= b
FT FT /gene= "AbsA2"
FT FT /product= "response regulator protein"

US5876987-A.
PN 02-MAR-1999.
PD 02-MAR-1999.
PF 06-FEB-1997; 796414.
PF 07-FEB-1996; US-011286.
PR 06-FEB-1997; US-796414.
PA (UNNS) UNIV MICHIGAN STATE.
PI Anderson TB, Brian P, Champness WC;
PI WPI: 99-189640/16.
DR P-PSDB; W78491, W78492.
DR Recombinant Streptomyces species - useful for the production of
PT recombinant proteins and the hyper-expression of antibiotics,
PT particularly actinorhodin and undecylprodigiosin
PS Claim 10; Fig 4; 30pp: English.
CC The sequences X21397-X21399 represent overlapping sequences from the AbsA
CC operon of Streptomyces coelicolor. This sequence corresponds to the
CC central fragment and contains the open reading frames for the Abs1 and
CC Abs2 genes, which encode a histidine kinase and response regulator for
CC the Abs1 gene, respectively. Inactivation of the Abs1 gene results in
CC mutation or insertion of heterologous DNA into the Abs1 gene results in
CC a Streptomyces organism which can hyperproduce heterologous proteins or
CC antibiotics, especially actinorhodin and undecylprodigiosin.
SO Sequence 2538 BP: 332 A; 901 C; 353 T;

	Query Match	4.9%	Score 38.8	DB 1	Length 2538
	Best Local Similarity	50.0%	Pred. No. 1.9		
	Matches 144	Conservative	0	Mismatches 122	Indels 2
				Gaps	
Qy	512	gtccaccagatgcctggagggcgcactgctcagcgaaagcgaccgcctgcga	571		
Db	1588	GCCTCCGAGGAGGCACGATGGCGGCGGAGCACCGCGGTATCAGGCGGCGCGCC	1529		
Qy	572	ggccatcgggatggactccgaaaggggcccacaagagctctagaagaggggacga	631		

Db 1528 GGGGCTCCCGCTGATCCGGCCCGGGTGGCGCAGCAGCGCGGCCCGGGAGCGTGGCGCC 1469
Qy 632 gagctctccctccagctgtcccccaaaagaccacttactgtatcctc--cagg 689
Db 1468 ACGGCCGGAGGGGGCGGCGAGGTCCGCGCGGGCGCGGCCGCCCTCTGTCGCGG 1409
Qy 690 cctctcggcagctgtgaggggggacccgggagcactctgtgagcccccatcagac 749
Db 1408 CCGTCGGGCCCCGGCGCGTGCCTGCCCGAGCCACCGCGCGTGCCTGCCGTCGG 1349
Qy 750 cctgcccc 757
Db 1348 CCGGCCCC 1341

RESULT 10
X24751/c
ID X24751 standard; cDNA; 7063 BP.
AC X24751;
DT 21-JUN-1999 (first entry)
DE Human interleukin-18 binding protein; IL-18BP; IL-18BPc; splice variant;
KW Interleukin-18 binding protein; IL-18BP; IL-18BPc; splice variant;
KW human; autoimmune disease; inflammation; diabetes; pancreatitis;
KW rheumatoid arthritis; graft rejection; sepsis; chronic hepatitis;
KW psoriasis; inflammatory bowel disease; multiple sclerosis;
KW ischemic heart disease; ischemic brain injury; gene therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 949..1542
FT FT /*tag= a
FT FT /product= "IL-18BPc"
FT FT 949..1032
FT FT /*tag= b
FT FT 1033..1539
FT FT /*tag= c

PN W09909063-A1.
PD 25-FEB-1999.
PR 13-AUG-1998; IL0379.
PR 22-JUL-1998; IL-125463.
PR 14-AUG-1997; IL-121554.
PR 27-AUG-1997; IL-121639.
PR 29-SEP-1997; IL-121860.
PR 06-NOV-1997; IL-122134.
PA (YEDA) YEDA RES & DEV CO LTD.
PI Dinarello C, Kim SH, Novick D, Rubinstein M;
DR WPI; 99-180975/15.
DR P-PSDB; W98006.
PT New interleukin-18 binding protein - useful for treating human
PT diseases, including autoimmune disease and inflammation
PS Claim 12; Page 55-58; 100pp; English.
CC The present sequence represents a cDNA clone coding for a human
CC interleukin-18 binding protein (IL-18BP) splice variant designated
CC IL-18BPc (see W98006). This is one of 4 novel splice variants (see
CC X24749-52), all coding for putative soluble proteins (see
CC W98004-07), that were detected following a screening of human
CC peripheral blood monocyte Jurkat T-cell, peripheral blood
CC mononuclear cell and spleen cell cDNA libraries using a probe
CC produced using IL-18BP-specific primers. IL-18BPc is a low
CC abundance splice variant. IL-18BP polypeptides capable of
CC binding IL-18 and/or modulating and/or blocking IL-18 activity are
CC provided. Methods for their isolation and recombinant production,
CC DNA vectors expressing them, vectors useful for their expression in
CC humans and other mammals, and antibodies against them are also
CC provided. IL-18BP polypeptides, and DNA encoding them, can be used
CC to treat conditions requiring the protein (claimed). Conditions
CC include autoimmune diseases, type I diabetes, rheumatoid arthritis,
CC graft rejections, inflammatory bowel disease, sepsis, multiple
CC sclerosis, ischemic heart diseases, ischemic brain injury, chronic
CC hepatitis, psoriasis, and chronic/acute pancreatitis. IL-18BP is
CC also useful for purifying IL-18 (claimed).
SQ Sequence 7063 BP; 1538 A; 1981 G; 1519 T;

Query Match 4.9%; Score 38.8; DB 1; Length 7063;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 152; Conservative 0; Mismatches 147; Indels 5; Gaps 2;

Qy 293 ccccatggaactccggggtccctccctgggaactacacaaaagaggaaccaggagcacca 352
Db 1486 CCTCTGTTCCTGGAGGCTCTCTCTTGGCTCTCCACCCAGAGCTGGGCCAGGACGCT 1427
Qy 353 gctggggaacaaaccctctccagccactccagatcagaaaggtaccagagtaggagga 412
Db 1426 GACGCTGGACAAACCTGTTTCAGGGTCCACGACACACAGGAAAGTTGGTGTGTCAGGG 1367
Qy 413 gaaggaggccctggtaccatccagaagccacgacgcttccacacagaactc---ca 469
Db 1366 CAGGGGTACGTGCT--CCAGCACCACAGGCTTGCACAGCTGCGTACCTGTCTCCACG 1309
Qy 470 tccccgggtggcctcttggtatcattaaagtccacggcgagggtcccccaccaggatgccct 529
Db 1308 TTCCCGGCTGGTGCCTCCCTCCACAGTCGGCTGGGAGGTGCTCAATGAAGGAACCAT 1249
Qy 530 ggaaggggccactggctcagcaggaagcagacacgcctcagccatcccggtggact 589
Db 1248 GCCCAGCAGTAGAGGATGCTGAAGTTGGGAAGCGCTGCAGGCCACACAGGATAAGCT 1189
Qy 590 ccgc 593
Db 1188 CAGC 1185

RESULT 11
ID Q62177 standard; cDNA; 2357 BP.
AC Q62177;
DT 10-SEP-1994 (first entry)
DE Human TLE-3 gene.
KW TLE-3; transducin-like enhancer of split protein; differentiation;
KW protein transport; cervix cancer; dysplasia; malignancy;
KW ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 22..2340
FT FT /*tag= a
FT PN W09408037-A.
PD 14-APR-1994.
PF 30-SEP-1993; U09333.
PR 30-SEP-1992; US-955011.
PA (MEDI-) MEDICAL RES COUNCIL.
PA (UYVA) UNIV YALE.
PI Artavanis-tsakonas S, Hill RE, Redhead NJ, Stifani S;
DR WPI; 94-135597/16.
DR P-PSDB; R51478.
PT New human transducin-like enhancers of split protein - and
PT associated multi-protein complexes, chimeric proteins,
PT antibodies, nucleic acid, etc., involved in nuclear-cytoplasmic
PT protein transport
PS Disclosure; Page 63-66; 112pp; English.
CC The nucleotide and deduced aa sequences of human transducin-like
CC enhancer of split proteins TLE-1 (Q62175, R51476), TLE-2 (Q62176,
CC R51477), TLE-3 (Q62177, R51478) and TLE-4 (Q62178, R51479) were
CC determined. The aa sequences were compared with that of Drosophila
CC E(spl)m9/10 protein (R51481). Comparison of the WD-40 domains of
CC these proteins defined the consensus residues shown in R51480. The
CC Cn motifs of the proteins were compared with those of the SV40 T
CC antigen, human c-myc, human p53, human A-myc and dorsal proteins
CC with respect to nuclear localization site, and casein-kinase II and
CC cdc2-kinase phosphorylation sites (sequences R51482-96). TLE can be
CC used to treat or diagnose (pre)neoplastic conditions, or
CC to study cell differentiation.
SQ Sequence 2357 BP; 548 A; 767 C; 646 G; 396 T;

Query Match 4.9%; Score 38.4; DB 1; Length 2357;
Best Local Similarity 46.9%; Pred. No. 2.3;

```
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 317 tgggaactaccacaagagagaaacacagagcaccagctgggggaacaacacacctctccag 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 TGGGAGATGACGCCGTCGCCCTTCCCTCCACAGAGGCCCTGGCAGGCCCGGCATCCCGAG 1445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 377 ccacctccagatgacaaagttaccacaggtgagagagaagagggccctgtaccatcca 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 GCACGCCGCGCAGATCAACACTCAGCCACGCGGGGGTGGTGTGTGCGGTGACCATCAG 1505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 437 gaagggccagcagcttcacacagaactccatccccgggtggccttcttggtatcataa 496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1506 CAACCCAGCAGGACGCTCTACACAGTGCAGAGGCGTGGTGAAGATCTGGGACATCAG 1565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 497 gctgccacgcgaggtctccaccagatgcccctggaggcgccactggctcagcgagaa 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1566 CCAGCCAGGCGAGAGAGGCCCTCTCCAGCTGGAGTGCCTGCAACAGGGACAATTACAT 1625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 557 ggcacacgcgctgcag 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1626 GCGTCTCTGCAAGCTG 1641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
Q45335
ID V80252 standard; cDNA; 2357 BP.
AC Q45335;
DT 10-SEP-1994 (first entry)
DE Human TLE-3 gene.
KW TLE-3; transducin-like enhancer of split protein; cell fate;
KW differentiation; cervix cancer; breast cancer; psoriasis; baldness;
KW ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT key 22..2340
FT cds /*tag= a
FT CDS
PN WO9407522-A.
PD 14-APR-1994.
PR 30-SEP-1993; U09339
PR 30-SEP-1992; US-954613.
PI (UYA) UNIV YALE.
PI Artavanis-tsakonak S, Stifani S;
DR WPI; 94-135221/16.
DR P-PSDB; R51111.
DR Transducin-like enhancer or split proteins and nucleic acids -
PT are for treatment of disorders of cell fate or differentiation
PT e.g. cervix cancer, breast cancer, psoriasis, baldness etc.
PS Disclosure; Page 96-99; 147pp; English.
CC The nucleotide and deduced aa sequences of human transducin-like
CC enhancer of split proteins TLE-1 (Q45333, R51109), TLE-2 (Q45334,
CC R51110), TLE-3 (Q45335, R51111) and TLE-4 (Q45336, R52953) were
CC determined. The aa sequences were compared with that of
CC Drosophila E(spl) m9/10 (R52955). Comparison of the WD-40 domains
CC of these proteins defined the consensus residues shown in R52954.
CC The Ccn motifs of the proteins were compared with those of SV40 T
CC antigen, human c-myc, human p53, human A-myp and dorsal protein with
CC respect to nuclear localization site, and casein kinase and cdk2-
CC kinase phosphorylation sites (sequences R52956-70).
CC
SQ Sequence 2357 BP; 548 A; 767 C; 646 G; 396 T;

Query Match 4.9%; Score 38.4; DB 1; Length 2357;
Best Local Similarity 46.9%; Pred. No. 2.3;
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 317 tgggaactaccacaagagagaaacacagagcaccagctgggggaacaacacacctctccag 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 TGGGAGATGACGCCGTCGCCCTTCCCTCCACAGAGGCCCTGGCAGGCCCGGCATCCCGAG 1445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 377 ccacctccagatgacaaagttaccacaggtgagagagaagagggccctgtaccatcca 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 GCACGCCGCGCAGATCAACACTCAGCCACGCGGGGGTGGTGTGTGCGGTGACCATCAG 1505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 317 tgggaactaccacaagagagaaacacagagcaccagctgggggaacaacacacctctccag 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 TGGGAGATGACGCCGTCGCCCTTCCCTCCACAGAGGCCCTGGCAGGCCCGGCATCCCGAG 1445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 377 ccacctccagatgacaaagttaccacaggtgagagagaagagggccctgtaccatcca 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 GCACGCCGCGCAGATCAACACTCAGCCACGCGGGGGTGGTGTGTGCGGTGACCATCAG 1505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 437 gaagggccacgacagcttccacacagaaactccatccccgggtggccttcttggtatcataa 496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1506 CAACCCAGCAGGACGCTCTACACAGTGCAGAGGCTGGTGAAGATCTGGGACATCAG 1565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 497 gctgccacgcgaggtctccaccagatgcccctggaggcgccactggctcagcgagaa 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1566 CCAGCCAGGCGAGCAGAGGCCCTCTCCAGCTGGAGTGCCTGCAACAGGGACAATTACAT 1625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 557 ggcacacgcgctgcag 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1626 GCGTCTCTGCAAGCTG 1641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
V80252
ID V80252 standard; cDNA; 470 BP.
AC V80252;
DT 01-MAR-1999 (first entry)
DE Modified Brazil nut 2S albumin gene AT2S1BN15 nucleotide sequence.
DE Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN19; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid; ds.
OS Bertholletia excelsa.
OS Synthetic.
FH Key Location/Qualifiers
FT key 3..452
FT cds /*tag= a
FT FT /gene= "AT2S1BN15"
FT FT /product= "modified Brazil nut 2S albumin seed
FT FT storage protein"
PN WO945458-A1.
PD 15-OCT-1998.
PR 06-APR-1998; U06673.
PR 08-APR-1997; US-042827.
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PI Gutteridge S;
DR WPI; 98-609902/51.
DR P-PSDB; W86272.
DR New modified Brazil Nut 2S albumin proteins - useful to provide
PT transgenic seed crops whose proteins have increased levels of
PT essential amino acids
PS Claim 5; Fig 10; 80pp; English.
CC The invention relates to modified Brazil Nut 2S albumin seed storage
CC proteins and nucleic acids encoding the modified proteins. The modified
CC Brazil nut 2S albumin proteins are rich in essential amino acids and the
CC encoding genes selected from a group comprising of BNCNSS, BN11, BN15,
CC BN17, BN18, BN19, BN153KW, AT2S1BN15, AT2S1BN19, AT2S1BN153W genes.
CC Chimeric genes comprising a nucleic acid fragment encoding a modified
CC Brazil nut 2S protein operably linked to suitable regulatory sequences,
CC integrated into the genome of a transformed plant can be used in
CC methods for increasing the essential amino acid content of seeds. The
CC products can be used to provide a human and animal dietary protein in
CC seed crop plants that is relatively high in essential amino acid content.
CC The present sequence represents a nucleotide sequence of a modified
CC Brazil Nut 2S albumin gene AT2S1BN15.
CC
SQ Sequence 470 BP; 138 A; 108 C; 142 G; 82 T;

Query Match 4.8%; Score 37.8; DB 1; Length 470;
Best Local Similarity 52.9%; Pred. No. 2.3;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 296 catggacttcggggccctcctgggaactaccacagaaggagaaaccagagcaccagct 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 CTGGAGTTCTGAGGAGGACCGACGCGAGATCATATGTCAGGAGGAGTGTAAAGAGCAGAT 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 ggggaaacaacacctctccggccacctccagatcgacaaaggtaccaggtgagagagaa 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 GCAGAAACAGAAAGATGTCAGCCCACTGCNAGATGTACATGAAACACAGCATGGAGGAG 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 ggaagccctgtaccatccagagagccacgga 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 CCCGTACAGAGCATGCCCAAGAGGAATGGA 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


us-09-215-435-118.rng

Wed May 24 09:19:46 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:55:49 ; Search time 203.6 seconds
(without alignments)
501.170 Million cell updates/sec

Title: US-09-215-435-118

Perfect score: 785

Sequence: 1 cggaaatccggggagtcgggtg.....tatggaataaagtcttttc 785

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/PCRTUS_COMB.seq.*
- 7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	5.5	7218	1	US-08-232-463-14
2	41	5.2	1137	7	Sequence 14, Appl Patent No. 5171843-8
3	39.4	5.0	697	7	Patent No. 5171843
4	39	5.0	1146	2	US-08-482-385A-1
5	39	5.0	2728	2	US-08-482-385A-5
6	38.8	4.9	2588	3	US-08-786-414B-6
7	36.6	4.7	2058	4	US-08-749-391-1
8	36.6	4.7	3177	1	US-08-042-747A-4
9	36.4	4.6	1124	4	US-08-966-316-11
10	36	4.6	2238	1	US-08-389-668A-1
11	36	4.6	2238	2	US-08-732-506-1
12	36	4.6	2238	6	PCT-US95-05768-1
13	36	4.6	2943	1	US-08-042-747A-7
14	35.8	4.6	1419	1	US-08-103-739B-1
15	35.8	4.6	1419	2	US-08-474-404-1
16	35.8	4.6	1419	3	US-08-485-845-1
17	35.8	4.6	1419	3	US-08-482-714-1
18	35.6	4.5	1072	1	US-07-971-096-1
19	35.6	4.5	1072	2	US-08-175-096-1
20	35.6	4.5	30001	1	US-08-125-468-1
21	35.6	4.5	30001	3	US-08-474-933-1
22	35.4	4.5	2214	5	US-08-864-038A-1
23	35.4	4.5	3331	5	US-08-864-038A-2
24	35.4	4.5	3331	5	US-08-864-038A-4
25	35	4.5	1505	1	US-07-915-246-1
26	35	4.5	2176	7	5320958-1
27	35	4.5	2793	2	US-08-209-747-1

Sequence 1, Appl Patent No. 5386025
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 17, Appl
Sequence 17, Appl
Sequence 17, Appl
Sequence 19, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 9, Appl
Sequence 9, Appl

US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pT7gpt-Fls
US-08-232-463-14

ALIGNMENTS

Query Match	5.2%;	Score 41;	DB 7;	Length 1137;
Best Local Similarity	47.5%;	Prod NO. 0.19;		
Matches 122;	Conservative	0;	Mismatches 135;	Indels 0;
Gaps	0;			
321	aactaccacaaagagagaaacaggagaccagccagctggggaaacacacacccctctccagccac	380		
yy				
362	aaccacagagatagacagctggacaaccagcaggagatagacagatagacagccag	421		
bb				
381	ctccagatgacaaggtacccaagatggaggagaaggagccctggtaccctccagaag	440		
yy				
422	caggagacagcagatggacagccagcgaggacagcagatggacaaccagcaggag	481		
bb				
441	gccacgacagcttccacacagaaactccatcccccggctggcctcttgatattaaagctg	500		
yy				
482	acagagcagatggacaaccagcagggtatagacagctggacaaccagcaggtgatagag	541		
bb				
501	ccacggcgagggtcccaccaggatgcccctggaggcgccactggtcagcagcagaagcga	560		
yy				
542	cagctggacaaccagcaggagatagacagcagatggacagccagcaggagatagacagctg	601		
bb				
561	caccgctcagccat	577		
yy				
602	gacgcccagcaggagat	618		
bb				

RESULT 3

```
;
;
; REGISTRATION NUMBER: 31,304
; REFERENCE/DOCKET NUMBER: PC8346C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-573-1189
; TELEFAX: 212-573-1939
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1146 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-482-385A-1

Query Match          5.0%; Score 39; DB 2; Length 1146;
Best Local Similarity 49.7%; Pred. No. 0.57;
Matches 99; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 250 tgaaggtaacctgcttcggtgcatagacagcttattctctgccccatggacttcggg 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 TGACCGAACCGGGCTGCTGACGAGAGCGCATCCGGGGCCGCCCGGAGGAGCGCGAGG 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 gctccctgggaactaccacaagaagagagaacacagagcaccagctggggaaacaacccc 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 977 CGATGCGCGCGGACCTGCGCGACGATGAACACGAGATCCGGCCCTGGACCCCATGGACC 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 ttccagccacctcccgatgacaaagtacccaggtgagaggaaggagccctggtac 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1037 TGTTGCCCATGTGTATGCCGAGCCACCCCGAGCTGCGGGAGCAGGAAGCCAGTTGC 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 ccattccagaagccacgga 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1097 GGGCCGAGCTGGCAGCGGA 1115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-08-482-385A-5
; Sequence 5, Application US/08482385A
; Patent No. 5728561
; GENERAL INFORMATION:
; APPLICANT: DENOYA,, CLAUDIO D.
; TITLE OF INVENTION: GENES ENCODING BRANCHED CHAIN ALPHA
; TITLE OF INVENTION: KETOACID DEHYDROGENASE FROM STREPTOMYCES SPECIES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETER C. RICHARDSON
; STREET: 235 EAST 42ND STREET, 20TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A
; ZIP: 10017-5755
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,385A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHEYKA,, ROBERT F.
; REGISTRATION NUMBER: 31,304
; REFERENCE/DOCKET NUMBER: PC8346C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-573-1189
; TELEFAX: 212-573-1939
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2728 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA

;
;
; REGISTRATION NUMBER: 31,304
; REFERENCE/DOCKET NUMBER: PC8346C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-573-1189
; TELEFAX: 212-573-1939
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1146 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-482-385A-5

Query Match          5.0%; Score 39; DB 2; Length 2728;
Best Local Similarity 49.7%; Pred. No. 0.66;
Matches 99; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 250 tgaaggtaacctgcttcggtgcatagacagcttattctctgccccatggacttcggg 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1319 TGACCGAACCGGGCTGCTGACGAGAGCGCATCCGGGGCCGCCCGGAGGAGCGCGAGG 1378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 gctccctgggaactaccacaagaagagagaacacagagcaccagctggggaaacaacccc 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1379 CGATGCGCGCGGACCTGCGCGACGATGAACACGAGATCCGGCCCTGGACCCCATGGACC 1438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 ttccagccacctcccgatgacaaagtacccaggtgagaggaaggagccctggtac 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1439 TGTTGCCCATGTGTATGCCGAGCCACCCCGAGCTGCGGGAGCAGGAAGCCAGTTGC 1498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 ccattccagaagccacgga 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1499 GGGCCGAGCTGGCAGCGGA 1517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-08-796-414B-6/c
; Sequence 6, Application US/08796414B
; Patent No. 5876987
; GENERAL INFORMATION:
; APPLICANT: Wendy C. Champness, Paul Brian
; APPLICANT: and Todd B. Anderson
; TITLE OF INVENTION: METHOD, DNA AND BACTERIA
; TITLE OF INVENTION: FOR HYPERPRODUCTION OF AN
; TITLE OF INVENTION: ANTIBIOTIC DUE TO
; TITLE OF INVENTION: DISRUPTION OF AN Absa
; TITLE OF INVENTION: GENE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch,
; COMPUTER: Acer
; OPERATING SYSTEM: MS-DOS 5.00
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/796,414B
; FILING DATE: February 6, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-297
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; TELEX: No. 5876987e
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2588
; TYPE: Nucleotide
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA
```


STREET: 112 East Pecan Street, Suite 2000
CITY: San Antonio
STATE: Texas
COUNTRY: USA
ZIP: 78205
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/042.747A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Raymond, W. Bradley
REGISTRATION NUMBER: 35186
REFERENCE/DOCKET NUMBER: S-0072.179
TELECOMMUNICATION INFORMATION:
TELEPHONE: 210-554-5500
TELEFAX: 210-226-8395
TELEX: 767609
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..249
FEATURE:
NAME/KEY: CDS
LOCATION: 1..249
US-08-042-747A-4

Query Match 4.7%; Score 36.6; DB 1; Length 3177;
Best Local Similarity 56.1%; Pred. No. 2.5; Mismatches 54; Indels 0; Gaps 0;
Matches 69; Conservative 0;
QY 401 caggatgaggagagggccctgtgtacccatccagaaagccagcgacaggtccacac 460
DB 1519 CGGCACGCACGTGAGGTGGCGCAGGTACGTACTACCTGCCAGCGCGCTTCTCAT 1578
QY 461 agaatccatcccccgggtggtcttctgtgatcattatgaagtcgcaagcgcgaggtccacca 520
DB 1579 CGCGTACCAGCCCTCCTCAGCAGCGGCTCGTGGAGCTGTACGTGGGAGCTCCTCCG 1638
QY 521 gga 523.
DB 1639 CGA 1641

RESULT 9
US-08-966-316-11
Sequence 11, Application US/08966316
Patent No. 5932445
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Au-Young, Janice
APPLICANT: Reddy, Roopa
APPLICANT: Murry, Lynn E.
APPLICANT: Mathur, Preete
TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA

ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/966.316
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0424 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LEUKNOT03
CLONE: 1880692
US-08-966-316-11

Query Match 4.6%; Score 36.4; DB 4; Length 1124;
Best Local Similarity 50.0%; Pred. No. 2.3;
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 26 gctgtgtctagcataaagcgagccagagagggcggtatgggagaagcctc 85
DB 20 GCGGTGTGGTGTAGTCCGTGGTGGCAGAGCGGAGCGGAGCTCTAGGGGTGGCAC 79
QY 86 cccactctcccccgcaagcgcgcatctgtgtctgtctctctctctctacccctgt 145
DB 80 CGGCCCCGAGAGGAGGATCGGGTCCGGATAGGGCTGCTGCTGTGTGCGGTGCT 139
QY 146 gatccccctccgtcgagctctctctctctctctctctctctctctctctctct 205
DB 140 GCTGAGCTTGGCTCGGCTCCTCGGATGAGAGGCGCAGGATGAATCCTTAGATTC 199
QY 206 ca 207
DB 200 CA 201

RESULT 10
US-08-389-668A-1
Sequence 1, Application US/08389668A
Patent No. 5637470
GENERAL INFORMATION:
APPLICANT: Kaczorowski, Gregory J.
APPLICANT: Garcia, Maria L.
APPLICANT: Leonard, Reid J.
APPLICANT: McManus, Owen B.
APPLICANT: Swanson, Richard J.
APPLICANT: Folander, Kimberly L.
TITLE OF INVENTION: NOVEL HETEROMULTIMER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Francis P. Bigley
STREET: 126 E. Lincoln Avenue, P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA

```

APPLICATION NUMBER: US/08/732,506
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05768
FILING DATE:
APPLICATION NUMBER: US 08/242,811
FILING DATE: 13-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bigley, Francis P.
REGISTRATION NUMBER: 36,356
REFERENCE/DOCKET NUMBER: 19226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-1249
TELEFAX: (908) 594-4270
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2238 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-732-506-1

Query Match 4.6% Score 36; DB 2; Length 2238;
Best Local Similarity 53.6%; Pred. No. 3.2;
Matches 75; Conservative 0; Mismatches 65; Indels 0; Gaps

QY 316 ctgggaattacacaaagagagagaccaggagaccagctggggaacaaacacctctctcca 375
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 CTGTGCTGTACACACAGGAGGACACGCGGACACAGAACACACAGTGTCTTACATCCCAA 422
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 376 gccacctccagatcacaaagtaccagagatggagagaaggagccctgggtaccatccc 435
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 CGAGCTGGACAACTACCAAGTGGCCCGCGGCGAGCTGGAGAAAGTCTAGAGCCAGGTTC 482
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 436 aagaagccacggacagcttc 455
|| || || ||||| |||||
Db 483 ACGAGAACGAGATTCTTC 502
|| || || ||||| |||||

RESULT 12
PCT-US95-05768-1
; Sequence 1, Application PC/TUS9505768
; GENERAL INFORMATION:
; APPLICANT: Kaczorowski, Gregory J.
; APPLICANT: Garcia, Maria L.
; APPLICANT: Leonard, Reid J.
; APPLICANT: McManus, Owen B.
; APPLICANT: Swanson, Richard J.
; APPLICANT: Folander, Kimberly L.
; TITLE OF INVENTION: NOVEL HETEROMULTIMER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Francis P. Bigley
; STREET: 126 E. Lincoln Avenue, P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05768
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,811
; FILING DATE: 13-MAY-1994

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:02:19 ; Search time 3550.87 Seconds
(without alignments)
-921.670 Million cell updates/sec

Title: US-09-215-435-118
Perfect score: 785
Sequence: 1 cggaaaccgggagtcggtg.....tatggaataaaagtcttttc 785

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues
Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA.*

1: /cgn2_6/ptodata/1/pna/PCUTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq.*

45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US095_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq.*
86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq.*
87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq.*
88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq.*
90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq.*
91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq.*
92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq.*
93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq.*
94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq.*
95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq.*
96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq.*
97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq.*
98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	785	100.0	785	43	US-09-215-435-118 Sequence 118, App
2	785	100.0	785	72	US-60-081-563-71 Sequence 71, Appl
3	770	98.1	865	1	PCT-US99-13418-68 Sequence 68, Appl

RESULT 2
US-60-081-563-71
; Sequence 71, Application US/60081563
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/081.563
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.027PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 785 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Testis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 72..704
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 72..161
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 13.2
; OTHER INFORMATION: seq LLLSLVLVPSAA/AP
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 772..777
; FEATURE:
; NAME/KEY: est
; LOCATION: 95..401
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 71..377
; OTHER INFORMATION: id AA397836
; FEATURE:
; NAME/KEY: est
; LOCATION: 61..103
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 38..80
; OTHER INFORMATION: id AA397836
US-60-081-563-71

Query Match 100.0%; Score 785; DB 72; Length 785;

Best Local Similarity 100.0%; Pred. No. 2.6e-157;
Matches 785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cggaaatccgggagtcgctgacccgggctgtggtctagcataaaagcgagccccagaaga 60
Db 1 CGGAATCCGGGAGTCCGGTGCACCCGGGCTGTGGTCTAGCATAAAGCGGAGCCCAAGA 60
Qy 61 agggcggggtatggagaaagcctccacactgcccccaagcgagcgcatctgctggtcc 120
Db 61 AGGGCGGGGTATGGAGAAAGCCTCCCACTGCCCCCGAAGCGGCATCTGCTGGTCC 120
Qy 121 tgcgtgctctctctctacccctgggtatccccctcgctgcagctcctatccatgatgctg 180
Db 121 TGCTGCTCTCTCTCTACCTGGTGTATCCCTCGCTGCAGCTCCTATCCATGATGCTG 180
Qy 181 acgccccaaagagagctccttgggtctcacaggcctccagagcctactccaaggctcagcc 240
Db 181 ACGCCCAAGAGAGCTCCTTGGGTCTCACAGGCTCCAGAGCTTCTCAAGGCTTTCAGCC 240
Qy 241 gactttcctgaaaggttaacctgcttcgggcatagacagcttattctgcccccatgg 300
Db 241 GACTTTCTCTGAAAGGTAACTCTGCTTCGGGGCATAGACAGCTTATCTCTGCCCCCATGG 300
Qy 301 acttcggggcctcctcctgggaactaccacaagagagaaccagagagaccagctgggga 360
Db 301 ACTTCCGGGGCTCCTCGGAACTACCAAAAGAGAGAGACCAGGACCACTAGCTGGGGA 360
Qy 361 acaacacccctcagcagccacctccagatcgacaaggttaccacaggtgagaggaagg 420
Db 361 ACAACACCTCTCCAGCCACCTTCCAGATCGACAGGTACCAAGGTACCCAGGATGGAGGA 420
Qy 421 ccctgtaccatccagaagggccagcagcttcacacagaaactccatcccggtgg 480
Db 421 CCTGTGATCCCATCAGAAAGGCCACGGACAGCTTCCACAGAACTCCATCCCGGGTGG 480
Qy 481 cctctggtatcattagctgccacgcgaggtccacacagagatccctggagggcgcc 540
Db 481 CCTTCTGGATCATTAAAGTGCACGCGGAGGTCCACAGGATGCCCTGGAGGGCGGCC 540
Qy 541 actggtcagcagagcagacaccgctcagggccatccggatggactccgcaaggga 600
Db 541 ACTGGCTCAGCGAGAGCGACACCCGCTCGAGGCCATCCGGGATGGACTCCCAAGGGA 600
Qy 601 cccacaagagctcctagaaggggagcagagctcctccactccaggtgtccccc 660
Db 601 CCCACAAGAGCTCTAGAAGAGGGGACCGAGAGCTCTCCACACTCCAGGCTGTCCCCC 660
Qy 661 gaaagaccacttactgtacatcctcaggccctctcggcagctgtagggtgggaccgg 720
Db 661 GAAAGACCCACTTACTGTACATCCTCAGGCCCTCTCGGCAGCTGTAGGGGTGGGACCGG 720
Qy 721 ggaagcactgctgtagcccccatcagaccctgcccccaagcaccatattggaataaagt 780
Db 721 GGAGCACCTGCTGTAGCCCCCATCAGACCTTGCCCCCAAGCACCATATGGAATAAAGTT 780
Qy 781 ctttc 785
Db 781 CTTTC 785

RESULT 3
PCT-US99-13418-68
; Sequence 68, Application PC/TUS9913418
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: 94 Human secreted proteins
; FILE REFERENCE: P2029PCT
; CURRENT APPLICATION NUMBER: PCT/US99/13418
; CURRENT FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16

Qy	721	gagcacctgctgtagcccccatcagacctgccccagaccataatggaatacaagtt	780
Db	786	ggagcacctgctgtagccccatcagacctgccccagaccataatggaatacaagtt	845
Qy	781	cttctc	785
Db	846	cttctc	850

```

RESULT 4
US-09-491-404-3685
; Sequence 3685, Application US/09491404
; GENERAL INFORMATION:
; APPLICANT: Tang, Yuanhua T.
; APPLICANT: Tillinghast, John
; APPLICANT: Sinku, Ankura
; APPLICANT: Liu, Chenghua
; APPLICANT: Drmanac, Radolje T.
; TITLE OF INVENTION: Novel Contigs Obtained
; FROM VARIOUS LIBRARIES
; FILE REFERENCE: 785
; CURRENT APPLICATION NUMBER: US/09/491.404
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 3796
; SOFTWARE: pt_sp_genes Version 1.0
; SEQ ID NO 3685
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (406)...(541)
; OTHER INFORMATION: this location contains
; OTHER INFORMATION: MGEASPPAPARHLLVLLLLLS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (406)...(801)
; OTHER INFORMATION: similar to g1517093.1
; OTHER INFORMATION: Run with FASTXY 3.3t00
US-09-491-404-3685

```

	Query Match	86.78;	Score 680.4;	DB 56;	Length 1232;
	Best Local Similarity	89.3%;	Pred. No. 4.2e-135;		
	Matches 784;	Conservative 0;	Mismatches 1;	Indels 93;	Gaps
Qy	1	cggaaatccgaggagtcgcgtgacccggcgctggctcgtacataaagcgagccagaaga	60		
Db	335	cggaaatccgaggagtcgcgtgacccggcgctggctcgtacataaagcgagccagaaga	394		
Qy	61	agggcgggggtatgggaagactcccacatgcgcccgcaaggcgcatctcgtgccc	120		
Db	395	agggcgggggtatgggaagactcccacatgcgcccgcaaggcgcatctcgtgccc	454		
Qy	121	tgtctgtctctcttacccttggtgatccccctccgctcgagctccctatccatgatgctg	180		
Db	455	tgtctgtctctcttacccttggtgatccccctccgctcgagctccctatccatgatgctg	514		
Qy	181	acgcccaagagagctcctttgggtctcaccaggcctccagagcctactccaagggtcttaagcc	240		
Db	515	acgcccaagagagctccttgggtctcaccaggcctccagagcctactccaagggtcttaagcc	574		
Qy	241	gacttttctgaaaagtaacctgcttcgggcatagacagctattctctgcgcccatgg	300		
Db	575	gacttttctgaaaagtaacctgcttcgggcatagacagctattctctgcgcccatgg	634		
Qy	301	acttcggggccctccctgggaactaccacaagaggagaacccaggagaccagctggggga	360		
Db	635	acttcggggccctccctgggaactaccacaagaggagaacccaggagaccagctggggga	694		
Qy	361	acaacacccctctcagccacctccagatcac-----	395		
Db	695	acaacacccctctcagccacctccagatcacgacacacagacagagagg	754		

```

1 EARLIER APPLICATION NUMBER: 60/089,510
2
3 EARLIER FILING DATE: 1998-06-16
4
5 EARLIER APPLICATION NUMBER: 60/089,509
6
7 EARLIER FILING DATE: 1998-06-16
8
9 EARLIER APPLICATION NUMBER: 60/090,112
10
11 EARLIER FILING DATE: 1998-06-22
12
13 EARLIER APPLICATION NUMBER: 60/090,113
14
15 EARLIER FILING DATE: 1998-06-22
16
17 NUMBER OF SEQ ID NOS: 502
18
19 SOFTWARE: PatentIn Ver. 2.0
20
21 SEQ ID NO 68
22
23 LENGTH: 865
24
25 TYPE: DNA
26
27 ORGANISM: Homo sapiens
28
29 FEATURE:
30
31 NAME/KEY: SITE
32
33 LOCATION: (445)
34
35 OTHER INFORMATION: n equals a t g, or c
36
37 PCT-US99-13418-68

```

Query Match	98.1%;	Score 770;	DB 1;	Length 865;
Best Local Similarity	99.1%;	Pred. No. 3.9e-154;		
Matches 778;	Conservative	5;	Mismatches 1;	Indels 1;
QY	1	c g a a t c c g g g a g t c c g g t a c c g g g c t g t g t c t a g c a t a a a g c g a g c c c a g a g a	60	
Db	67	c g g a a t c c c g g a g t c c g g t a c c g g g c t g t g t c t a g c a t a a a g c g a - s c c a g a g a	125	
QY	61	a g g g g c g g g g t a t g g g a a a g c t c c c a c a c t c c c c c c g c a a g c g g c a t c t g t g t c c	120	
Db	126	a g g g c g g g g t a t g g g a a a g c t c c c a c a c t g c c c c g a a g c g g c a t c t g t g t c c	185	
QY	121	t g t g t g t c c t c t a c c t g g t a t c c c t c c g t c a g t c g a g t c c t a t c c a t g a t g t g	180	
Db	186	t g t g t g t c c t c t a c c t g g t a t c c c t c c g t c a g t c c t a t c c a t g a t g t g	245	
QY	181	a c g c c c a a g a g a c t c c t t g g t c t a c a g g c c t c c a g c t a c t c c a a g g t c c a g c	240	
Db	246	a c g c c a a g a g a c t c c t t g g t c t a c a g g c c t c c a g c t a c t c c a a g g t t c a g c	305	
QY	241	g a c t t t c c t g a a a g g t a a c c t g t t c g g g g c a t a g a c a g t t a t t c t g c c c c a t g g	300	
Db	306	g a c t t t c c t g a a a g g t a a c c t g t t c g g g g c a t a g a c a g t t a t t c t g c c c c a t g g	365	
QY	301	a c t t c g g g g c c t c c t g g g a a c t a c c a a g a g g a a c c a g g a g a c c a g a c a g t g g g a	360	
Db	366	a c t t c g g g g c c t c c t g g g a a c t a c c a a g a g g a a c c a a g a g a c c a r c t g g g a	425	
QY	361	a a a c a c c c t c c a g c c a c t c c a g t c g a c a a g t a c c c a g a t g g a g g a a a g g a g g	420	
Db	426	a a a c a c c y t c m a g c a n t c a g a t c g a a g g t a c c c a g a t g g a g g a a a g g a g g	485	
QY	421	c c c t g t a c c a t c c a a a g g c a c a g a g t t c c a c a g a a c t c c a t c c c c g g t g g	480	
Db	486	c c c t g t a c c a t c c a a a g g c a c a g a g t t c c a c a g a a c t c c a t c c c c g g t g g	545	
QY	481	c c t t c t g a t a t t a a g t c c a c g c g a g g t c c c a c a g g a t c c c t g a g g g c g c c	540	
Db	546	c c t t c t g a t a t t a a g t c c a c g c g a g g t c c c a c a g g a t c c c t g a g g a g c g c c	605	
QY	541	a t g t g c t a g c g a a g c g a c a c c g t c a g c c a t c c g a g a t g g a t c c g c a a g g g a	600	
Db	606	a t g t g c t a g c g a a g c g a c c c g t c g a g g c a t c c g g a t g g a c t c c g a a g g g a	665	
QY	601	c c c a a a g a g c t c t a g a a g g g a c c a g a g c t c t c c c a c t c a g g t g t c c c c c c	660	
Db	666	c c c a a a g a g c t c t a g a a g g g a c c a g a g c t c t c c c a c t c a g g t g t c c c c c c	725	
QY	661	g a a a c c c a c t a c t a c a t c a g c c c t c g g a g c t g t a g g g t g g g a c c g g	720	
Db	726	g a a a c c c a c t a c t a c a t c c t c g g c c t c g g a g c t g t a g g g t g g g a c c g g	785	


```
QY 393 ----- 393
Db 755 tgctgatctcgcagaatgtgtgcatccattcaaccacgagggaggttcgaggtg 814
QY 393 -----aaggtaccacaggtatgagagagaaggagccctgtgtaccatccagaagccacgg 447
Db 815 atttgaaggtaccacaggtatgagagagaaggagccctgtgtaccatccagaagccacgg 874
QY 448 acagttccacacagaactccatcccggtggccttctgtgatacttaagtgcacagc 507
Db 875 acagttccacacagaactccatcccggtggccttctgtgatacttaagtgcacagc 934
QY 508 ggaggtccacacaggtatccctgagggcgccactggtcagcgagagcgacacgcc 567
Db 935 ggaggtccacacaggtatccctgagggcgccactggtcagcgagagcgacacgcc 994
QY 568 tgcaggccatccggatggatccgcgaaggggacccacaaaggaagtcctagaagggga 627
Db 995 tgcaggccatccggatggatccgcgaaggggacccacaaaggaagtcctagaagggga 1054
QY 628 ccgagagatctccacatccaggtgtccccccgaaagacccacttactgtacatctca 687
Db 1055 ccgagagatctccacatccaggtgtccccccgaaagacccacttactgtacatctca 1114
QY 688 ggcctctcgcagctgtaggggtgaggaccgggagcacctgctgtagcccccatcag 747
Db 1115 ggcctctcgcagctgtaggggtgaggaccgggagcacctgctgtagcccccatcag 1174
QY 748 accctgccccaaagcaccatattgaaataaagtctttc 785
Db 1175 accctgccccaaagcaccatattgaaataaagtctttc 1212
```

RESULT 5

```
US-09-215-435-311
; Sequence 311, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215.435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 311
; LENGTH: 916
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..815
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..179
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 13.199998092651
; OTHER INFORMATION: seq LLLSTLVPSAA/AP
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 883..888
; FEATURE:
; NAME/KEY: polyA_site
```

```
; LOCATION: 905..916
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 487,661
; OTHER INFORMATION: n=a, g, c or t
US-09-215-435-311

Query Match      84.3%, Score 661.4, DB 43, Length 916;
Best Local Similarity 87.2%, Pred. No. 4.5e-131;
Matches 766; Conservative 17; Mismatches 1; Indels 94; Gaps 2;

QY 1  cggaaatccggagtcgggtgacccgggtgtgtgttagcataaaagcgagccagaaga 60
Db 20  cggaaatccggagtcgggtgacccgggtgtgtgttagcataaaagcgag-cagaaga 78
QY 61  agggcggggtatgggagaagcctccacatgcccccgcaaggcggaatctgtcgtgtcc 120
Db 79  agggcggggtatgggagaagcctccacatgcccccgcaaggcggaatctgtcgtgtcc 138
QY 121  tgcgtcgtcctctctaccctggtgatccctccgctgcagctcctatccatgatgtg 180
Db 139  tgcgtcgtcctctctaccctggtgatccctccgctgcagctcctatccatgatgtg 198
QY 181  acgcccagaagagctccttgggtctcacaggcctccagagcctcagagcctcacaagcttcagcc 240
Db 199  acgcccagaagagctccttgggtctcacaggcctccagagcctcagagcctcacaagcttcagcc 258
QY 241  gactttctcgaaggttaacctgttcggggcatagacagcttattctctgcccccatgg 300
Db 259  gactttctcgaaggttaacctgttcggggcatagacagcttattctctgcccccatgg 318
QY 301  acttcggggcctcctcgggaactaccacaaagagagaaaccagagagacacagctcgggga 360
Db 319  acttcggggcctcctcgggaactaccacaaagagagaaaccagagagacacagctcgggga 378
QY 361  acaacacccctcctcagccacctccagatcgac----- 393
Db 379  acaacacccctcctcagccacctccagatcgacaaagatgacogacaacaaagacagagagg 438
QY 393 ----- 393
Db 439  tgctgatctcgcagaatgtgtggtgcattcaaccavcggagggttcgaggggtg 498
QY 393 -----aaggtaccacaggtatgagagagaaggagccctgtgtaccatccagaagccacgg 447
Db 499  atttgaaggtgccaggtatggagaaaggagggcctgtgtaccmtccaraaggccacgg 558
QY 448  acagcttcacacagaactccatccccgggtgtgcttcttgatcattagctgcacagc 507
Db 559  acagcttcacacagaactccatccccgggtgtgcttcttgatcattagctgcacagc 618
QY 508  ggaggtccacacaggtatgccttgaggggcgccactggtcagcgagagcgacacccgcc 567
Db 619  ggaggtccacacaggtatgccttgaggggcgccactggtcancgaraagcgacacccgcc 678
QY 568  tgcaggccatccgggtatggactccgcaagggggacccacaaagagacgtccttagaaggggga 627
Db 679  tgcaggccatccgggtatggactccgcaagggggacccacaaagagacgtccttagaaggggga 738
QY 628  ccgagagctcctccacatccaggtgtccccccgaaagacccacttactgtacatctca 687
Db 739  ccgagagctcctccacatccaggtgtccccccgaaarammccacttactgtacatctca 798
QY 688  ggcctctcgcagctgtaggggtgggaccgggagcacctgcctgtagcccccatcag 747
Db 799  ggcctctcgcagctgtaggggtgggaccgggarmacctgctgtagcccccatcar 858
QY 748  accctgccccaaagcaccatattgaaataaagtctttc 785
Db 859  accctgccccaaagcaccatattgaaataaagtctttc 896
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	84.3%	Score	661.4	DB	70	Length	916
Best Local Similarity	87.2%	Pred.	No. 4.5e-131				
Matches	766	Conservative	17	Mismatches	1	Indels	94
Gaps							

QY	1	cggaatccgggagtcggtgaccgggctggtgctagacataaagcgagcgacgaaga	60
DB	20	CGGAATCCGGGAGTCGGGTGACCCGGGCTG7GGTCTAGCATTAAGCGGAG-CCAGAAGA	78
QY	61	agggcggggtatggagaagctcccacacctgcccccgcaagcgcgatctgctggtcc	120
DB	79	AGGGCGGGGTATGGAGAGGCTCCCCACCTGCGCCCGCAAGCGGCATCTGCTGGTCC	138
QY	121	tgtgtgtctctctctacccctggtgatccctccgctcgagctccctatccatgatgctg	180
DB	139	TGCTGTGCTCTCTCTACCTCGGTGGTGATCCCTCGCTGCGTGCAGTCTCTATCCATGATGCTG	198
QY	181	acgccaagagagctcttgggtctcacaggctccagagcctactccaagggttcagcc	240
DB	199	ACGCCAAGAGAGTCTCTGGGTCTCACAGGCTCCAGAGCTTACTCAAGGCTTCAGCC	258
QY	241	gactttctgaaagtaacctgtcttcggggcatagacagcttattcttgcgcccatgg	300
DB	259	GACTTTTCTGAAAGTAACTCTCTCGGGGCATAGACAGCTTATCTCTGCCCCCATGG	318
QY	301	acttccgggctccctgggaactccacaaagagagaaacacagagcagcaccagctggga	360
DB	319	ACTTCGGGGCTCCCTGGGACTACCAAAAGAGGAGAACACGAGCAGCCAGCTGGGGA	378
QY	361	acacacccctctcagaccactccaatcgac-----	393
DB	379	ACAACACCTCTCCAGGCACCTCCAGATCGACAAAGATGACCCGACAAAGACAGAGAGG	438
QY	393	-----	393
DB	439	TGCTGATCTCCGAGAACTGTGTGGCATCTTCAACCAVCGGAGGGGANTTCGAGGGTG	498
QY	393	-----aagttaccagatgagagagagagcgctggtaccatccagaaggccacgg	447
DB	499	ATTTGAAGTGTGCCAGAGTGGAGGARAAGNGGCCCTGGTACCCMTCCARAAGGCCACGG	558
QY	448	acagcttcacacagaactccatccccgggtggcctctggtatcatctaagctgccacgc	507
DB	559	ACAGCTTCCACACAGAACTCCATCCGGGTGGCCTTCTGGATCATTAAGCTGCCACGCG	618
QY	508	ggaggtccacacagatccctggagggcgccactggctcagcgagagagcgacaccgcc	567
DB	619	GGAGGTGCCACGAGATGCCCTGGAGGGCGCCACTGGCTCANCAGARAAGCGACACCGCC	678
QY	568	tgcaggccatccggatggaactccgcaagggggacccacagagcgctctagaagaggga	627
DB	679	TGCAGGCCATCCGGGTGAC'TCCCAAGGGAGCCCAAGGACRTCTTDAARAGGGGA	738
QY	628	ccgagagctctccactccagctgtccccccgaagaccacttactgtatcactca	687
DB	739	CCGARAGCTCTCCCACTCCAGCTTCCCGCGAAARAMMACTTACTGTACATCTCA	798
QY	688	ggcctctcgagctgtagggtgggacccggggagcacctgctagcccccatcag	747
DB	799	RGCCCTCTCGCAGCTGTARGGTGGGGACCGGGARMACCTGCTGTAGCCCCCATCAR	858
QY	748	acctgcccacacccatatgaaataaagtcttc	785
DB	859	ACCTTGCCCCAAGCACCCATATGGAATTAAGTTCTTTC	896

RESULT	7
US-09-009-325-13	
; Sequence 13, Application	US/0909325A
; GENERAL INFORMATION:	

FILE REFERENCE: MEI-008DV3
CURRENT APPLICATION NUMBER: US/09/009,325A
CURRENT FILING DATE: 1998-01-20
EARLIER APPLICATION NUMBER: 08/842,898
EARLIER FILING DATE: 1997-04-17
EARLIER APPLICATION NUMBER: 60/071,589
EARLIER FILING DATE: 1998-01-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 928
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (75)..(800)
US-09-009-325-13

Query Match 84.1%; Score 660; DB 35; Length 928;
Best Local Similarity 89.1%; Pred. No. 8.9e-131;
Matches 763; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
Qy 23 ccgggctgtgtctagcataaagcgagccagcaagaagaagggcggtatggagaagc 82
Db 26 ccgggctgtgtctagcataaagcgagccagcaagaagaagggcggtatggagaagc 85
Qy 83 ctccccactgcccccgcaagcgccatctgtgtctgtctgtctgtctctctaccc 142
Db 86 ctccccactgcccccgcaagcgccatctgtgtctgtctgtctctctctaccc 145
Qy 143 ggtatccctccgctcagctctctccatgctgacgcccagagagctctctggg 202
Db 146 ggtatccctccgctcagctctctccatgctgacgcccagagagctctctggg 205
Qy 203 tctcacaggctccagagctctctccatgctgacgcccagagagctctctggg 262
Db 206 tctcacaggctccagagctctctccatgctgacgcccagagagctctctggg 265
Qy 263 gcttcggggatagacagcttattcttgcgccatgacttcggggcctctgggaa 322
Db 266 gcttcggggatagacagcttattcttgcgccatgacttcggggcctctgggaa 325
Qy 323 ctaccacaaagagagaaccagagcaccagctgggggaacacaccctctccagccact 382
Db 326 ctaccacaaagagagaaccagagcaccagctgggggaacacaccctctccagccact 385
Qy 383 ccagatcgac----- 393
Db 386 ccagatcgacagatgaccgacacaaagagagagtgctgatctccgagaatgtggt 445
Qy 393 -----aagttaccaggatgga 409
Db 446 ggcattccattcaacagcgaggggagcttcgaggtgatttgaaggtaaccagatgga 505
Qy 410 ggaagaggccctgtgtaccctccagaaagccagagacttccacagactcca 469
Db 506 ggaagaggccctgtgtaccctccagaaagccagagacttccacagactcca 565
Qy 470 tccccgggtgctcttgatcattaaagctccagcgcgagggtccacaccaggtgcct 529
Db 566 tccccgggtgctcttgatcattaaagctccagcgcgagggtccacaccaggtgcct 625
Qy 530 ggaaggcgccactggctcagcgaagcgacacgcctcgagggccatccgggatggact 589
Db 626 ggaaggcgccactggctcagcgaagcgacacgcctcgagggccatccgggatggact 685
Qy 590 ccgcaaggggccacacagagcgtccctagaagaggagccagagctcctccactccag 649
Db 686 ccgcaaggggccacacagagcgtccctagaagaggagccagagctcctccactccag 745
Qy 650 gctgtccccccgaaagacccacttactgtacatctcagggccctctcggaagctgtagg 709
|||||

Db 746 gctgtccccccgaaagacccacttactgtacatctcagggccctctcggcagctgtagg 805
Qy 710 gtggggaccggggagagcacttctgttagcccccatcagacccctgcccccaagcaccatag 769
Db 806 gtggggaccggggagagcacttctgttagcccccatcagacccctgcccccaagcaccatag 865
Qy 770 gaaataaagtctcttc 785
Db 866 gaaataaagtctcttc 881

RESULT 8
US-09-009-685-13
Sequence 13, Application US/09009685A
GENERAL INFORMATION:
APPLICANT: McCarthy, Sean A.
TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
FILE REFERENCE: MEI-008DV2
CURRENT APPLICATION NUMBER: US/09/009,685A
CURRENT FILING DATE: 1998-01-20
EARLIER APPLICATION NUMBER: 08/843,704
EARLIER FILING DATE: 1997-04-16
EARLIER APPLICATION NUMBER: 08/842,898
EARLIER FILING DATE: 1997-04-17
EARLIER APPLICATION NUMBER: 60/071,589
EARLIER FILING DATE: 1998-01-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 928
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (75)..(800)
US-09-009-685-13

Query Match 84.1%; Score 660; DB 35; Length 928;
Best Local Similarity 89.1%; Pred. No. 8.9e-131;
Matches 763; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
Qy 23 ccgggctgtgtctagcataaagcgagccagcaagaagaagggcggtatggagaagc 82
Db 26 ccgggctgtgtctagcataaagcgagccagcaagaagaagggcggtatggagaagc 85
Qy 83 ctccccactgcccccgcaagcgccatctgtgtctgtctgtctgtctctctaccc 142
Db 86 ctccccactgcccccgcaagcgccatctgtgtctgtctgtctctctctaccc 145
Qy 143 ggtatccctccgctcagctctctccatgctgacgcccagagagctctctggg 202
Db 146 ggtatccctccgctcagctctctccatgctgacgcccagagagctctctggg 205
Qy 203 tctcacaggctccagagctctctccatgctgacgcccagagagctctctggg 262
Db 206 tctcacaggctccagagctctctccatgctgacgcccagagagctctctggg 265
Qy 263 gcttcggggatagacagcttattcttgcgccatgacttcggggcctctgggaa 322
Db 266 gcttcggggatagacagcttattcttgcgccatgacttcggggcctctgggaa 325
Qy 323 ctaccacaaagagagaaccagagcaccagctgggggaacacaccctctccagccact 382
Db 326 ctaccacaaagagagaaccagagcaccagctgggggaacacaccctctccagccact 385
Qy 383 ccagatcgac----- 393
Db 386 ccagatcgacagatgaccgacacaaagagagagtgctgatctccgagaatgtggt 445
Qy 393 -----aaggtaccaggatgga 409
|||||

Db 446 ggcattcattcaaccagcggaggagcttcagaggtgattgaagggtaccaggatgga 505
Qy 410 ggagaaaggagccctggtaccattccagaagccagcagcagctccacacagaactcca 469
Db 506 ggagaaaggagccctggtaccattccagaagccagcagcagctccacacagaactcca 565
Qy 470 tccccgggtgcttctgatacttaagctgccacggcgaggtccaccaggatgccct 529
Db 566 tccccgggtgcttctgatacttaagctgccacggcgaggtccaccaggatgccct 625
Qy 530 ggagggcgccactggtcagcagagcagacaccgcctgcaggccatccggatggact 589
Db 626 ggagggcgccactggtcagcagagcagacaccgcctgcaggccatccggatggact 685
Qy 590 ccgcaaggaggacccacaaagacgtctagagagggagacagctctccactccag 649
Db 686 ccgcaaggaggacccacaaagacgtctagagagggagacagctctccactccag 745
Qy 650 gctgtccccccgaaagaccacttactgtatctatctcctcagccctctcgcagctgtagg 709
Db 746 gctgtccccccgaaagaccacttactgtatctatctcctcagccctctcgcagctgtagg 805
Qy 710 gtggggaccgggagacactcctgtgtagcccccacacagacccctgcacacagccatag 769
Db 806 gtggggaccgggagacactcctgtgtagcccccacacagacccctgcacacagccatag 865
Qy 770 gaaataaaagtctttc 785
Db 866 gaaataaaagtctttc 881

RESULT 9
US-09-009-706-13
; Sequence 13, Application US/09009706A
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-008DV1
; CURRENT APPLICATION NUMBER: US/09/009,706A
; EARLIER FILING DATE: 1998-01-20
; EARLIER FILING DATE: 1997-04-17
; EARLIER FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(800)
US-09-009-706-13

Query Match 84.1%; Score 660; DB 35; Length 928;
Best Local Similarity 89.1%; Pred. No. 8,9e-131;
Matches 763; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

Qy 23 ccgggctgtgtctagcataaaagcggagcccgagaggggggggggtatgggagaagc 82
Db 26 ccgggctgtgtctagcataaaagcggagcccgagaggggggggtatgggagaagc 85
Qy 83 ctccccactgccccccgcaagcgccatctggtctctgtctgtctctctctaccct 142
Db 86 ctccccactgccccccgcaagcgccatctggtctctgtctgtctctctctaccct 145
Qy 143 ggtgatccccctccgtgcagctctctatccatgatctgtagcgcacagagagctccttggg 202
Db 146 ggtgatccccctccgtgcagctctctatccatgatctgtagcgcacagagagctccttggg 205

Qy 203 tctcacagcctccagagcctactccaaggttcagcagcagcttttctgaaaggtaacct 262
Db 206 tctcacagcctccagagcctactccaaggttcagcagcagcttttctgaaaggtaacct 265
Qy 263 gcttcggggcatagacagcttattctctgcccccatggacttcgggggctccctcgggaa 322
Db 266 gcttcggggcatagacagcttattctctgcccccatggacttcgggggctccctcgggaa 325
Qy 323 ctaccacaaagagagaaacagcagcagctggggaaacacacacctctccagcaccct 382
Db 326 ctaccacaaagagagaaacagcagcagctggggaaacacacacctctccagcaccct 385
Qy 383 ccagatgcac----- 393
Db 386 ccagatgcacaagatgaccgacacaaagacagagaggtgctgatctccgagaagtgtgt 445
Qy 393 -----aaggtaccacaggtgga 409
Db 446 ggcattcattcaaccagcggagggtctcgaggggtgattgaagggtaccaggatgga 505
Qy 410 ggagaaaggagccctggtaccattccagaagccagcagcagcttccacacagaactcca 469
Db 506 ggagaaaggagccctggtaccattccagaagccagcagcagcttccacacagaactcca 565
Qy 470 tccccgggtgcttctgatacttaagctgccacggcgaggtccaccaggatgccct 529
Db 566 tccccgggtgcttctgatacttaagctgccacggcgaggtccaccaggatgccct 625
Qy 530 ggagggcgccactggtcagcagagcagacccgcctgcaggccatccggatggact 589
Db 626 ggagggcgccactggtcagcagagcagacaccgcctgcaggccatccggatggact 685
Qy 590 ccgcaaggaggacccacaaagacgtcttagaagagggggagcagagctctccactccag 649
Db 686 ccgcaaggaggacccacaaagacgtcttagaagagggggagcagagctctccactccag 745
Qy 650 gctgtccccccgaaagaccacttactgtatctatctcctcagccctctcgcagctgtagg 709
Db 746 gctgtccccccgaaagaccacttactgtatctatctcctcagccctctcgcagctgtagg 805
Qy 710 gtggggaccgggagacactcctgtgtagcccccacacagacccctgcacacagccatag 769
Db 806 gtggggaccgggagacactcctgtgtagcccccacacagacccctgcacacagccatag 865
Qy 770 gaaataaaagtctttc 785
Db 866 gaaataaaagtctttc 881

RESULT 10
US-09-009-802-13
; Sequence 13, Application US/09009802B
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-008
; CURRENT APPLICATION NUMBER: US/09/009,802B
; EARLIER FILING DATE: 1998-01-20
; EARLIER FILING DATE: 1997-04-17
; EARLIER FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(800)
US-09-009-802-13

```
Query Match      84.1%; Score 660; DB 35; Length 928;
Best Local Similarity 89.1%; Pred. No. 8.9e-131;
Matches 763; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

Qy 23 ccgggctgtgtctagcataaagcgagccagaaagggcggtgtatggagaagc 82
    |||||
Db 26 ccgggctgtgtctagcataaagcgagccagaaagggcggtgtatggagaagc 85

Qy 83 ctccccacctgcccccgcaagggcgatctgtgtgtctgtgtctctctctaccc 142
    |||||
Db 86 ctccccacctgcccccgcaagggcgatctgtgtgtctgtgtctctctctaccc 145

Qy 143 ggtgatccccctccgctgagctctatccatgatgtgacgcccaagagagctcctggg 202
    |||||
Db 146 ggtgatccccctccgctgagctctatccatgatgtgacgcccaagagagctcctggg 205

Qy 203 tctcacaggctccagagcttactccaaaggcttcagccgacttttctgaaaggtaacct 262
    |||||
Db 206 tctcacaggctccagagcttactccaaaggcttcagccgacttttctgaaaggtaacct 265

Qy 263 gcttcgggggataagacagcttattctgtgccccatggaacttcggggcctccctgggaa 322
    |||||
Db 266 gcttcgggggataagacagcttattctgtgccccatggaacttcggggcctccctgggaa 325

Qy 323 ctaccacaaagagagaccagagcaccagctggtgggaacacacccctctccagccacct 382
    |||||
Db 326 ctaccacaaagagagaccagagcaccagctggtgggaacacacccctctccagccacct 385

Qy 383 ccagatcgac----- 393
    |||||

Db 386 ccagatcgacaaagatgaccgacaaacagagaggtgctgatctccgagaatgtggt 445

Qy 393 -----aagtaaccaggatgga 409
    |||||

Db 446 ggcattccattcaaccaggaggagcttcctgaggtgattgaaggtaccagagtgga 505

Qy 410 ggagaagaggccctgttaccatccagaagccacgagacgttccacacagaactcca 469
    |||||
Db 506 ggagaagaggccctgttaccatccagaagccacgagacgttccacacagaactcca 565

Qy 470 tccccgggtgacctcttgatcattaaagtccacggcgagggtcccccaggatgcacct 529
    |||||
Db 566 tccccgggtgacctcttgatcattaaagtccacggcgagggtcccccaggatgcacct 625

Qy 530 ggaggggcgccactggtctagcaggaagcagacccgctgcaggccatccgggatgacct 589
    |||||
Db 626 ggaggggcgccactggtctagcaggaagcagacccgctgcaggccatccgggatgacct 685

Qy 590 ccgcaaggggacccacaaagagcttcttagaagaggacacccgctgcaggccatccggtgacct 649
    |||||
Db 686 ccgcaaggggacccacaaagagcttcttagaagaggacacccgctgcaggccatccggtgacct 745

Qy 650 gctgtccccccgaaagaccacttactgtatcatctcagggccctctcggcagctgtaggg 709
    |||||
Db 746 gctgtccccccgaaagaccacttactgtatcatctcagggccctctcggcagctgtaggg 805

Qy 710 gtggggaccggggagacctgctgtagcccccatcagacctgcccaagcaccatag 769
    |||||
Db 806 gtggggaccggggagacctgctgtagcccccatcagacctgcccaagcaccatag 865

Qy 770 gaaataaagtcttttc 785
    |||||
Db 866 gaaataaagtcttttc 881
    |||||
```

RESULT 11

```
US-09-469-604-13
; Sequence 13, Application US/09469604
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
```

```
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: MEI-008DV4
; CURRENT APPLICATION NUMBER: US/09/469,604
; CURRENT FILING DATE: 1999-12-22
; EARLIER APPLICATION NUMBER: 08/843,704
; EARLIER FILING DATE: 1997-04-16
; EARLIER APPLICATION NUMBER: 08/842,898
; EARLIER FILING DATE: 1997-04-17
; EARLIER APPLICATION NUMBER: 60/071,589
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 09/009,325
; EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(800)
US-09-469-604-13

Query Match      84.1%; Score 660; DB 56; Length 928;
Best Local Similarity 89.1%; Pred. No. 8.9e-131;
Matches 763; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

Qy 23 ccgggctgtgtctagcataaagcgagccagaaagggcggtgtatggagaagc 82
    |||||
Db 26 ccgggctgtgtctagcataaagcgagccagaaagggcggtgtatggagaagc 85

Qy 83 ctccccacctgcccccgcaagggcgatctgtgtgtctgtgtctctctctaccc 142
    |||||
Db 86 ctccccacctgcccccgcaagggcgatctgtgtgtctgtgtctctctctaccc 145

Qy 143 ggtgatccccctccgctgagctctatccatgatgtgacgcccaagagagctcctggg 202
    |||||
Db 146 ggtgatccccctccgctgagctctatccatgatgtgacgcccaagagagctcctggg 205

Qy 203 tctcacaggctccagagcttactccaaaggcttcagccgacttttctgaaaggtaacct 262
    |||||
Db 206 tctcacaggctccagagcttactccaaaggcttcagccgacttttctgaaaggtaacct 265

Qy 263 gcttcgggggataagacagcttattctgtgccccatggaacttcggggcctccctgggaa 322
    |||||
Db 266 gcttcgggggataagacagcttattctgtgccccatggaacttcggggcctccctgggaa 325

Qy 323 ctaccacaaagagagaccagagcaccagctggtgggaacacacccctctccagccacct 382
    |||||
Db 326 ctaccacaaagagagaccagagcaccagctggtgggaacacacccctctccagccacct 385

Qy 383 ccagatcgac----- 393
    |||||

Db 386 ccagatcgacaaagatgaccgacaaacagagaggtgctgatctccgagaatgtggt 445

Qy 393 -----aagtaaccaggatgga 409
    |||||

Db 446 ggcattccattcaaccaggaggagcttcctgaggtgattgaaggtaccagagtgga 505

Qy 410 ggagaagaggccctgttaccatccagaagccacgagacgttccacacagaactcca 469
    |||||
Db 506 ggagaagaggccctgttaccatccagaagccacgagacgttccacacagaactcca 565

Qy 470 tccccgggtgacctcttgatcattaaagtccacggcgagggtcccccaggatgcacct 529
    |||||
Db 566 tccccgggtgacctcttgatcattaaagtccacggcgagggtcccccaggatgcacct 625

Qy 530 ggaggggcgccactggtctagcaggaagcagacccgctgcaggccatccgggatgacct 589
    |||||
Db 626 ggaggggcgccactggtctagcaggaagcagacccgctgcaggccatccgggatgacct 685

Qy 590 ccgcaaggggacccacaaagagcttcttagaagaggacacccgctgcaggccatccggtgacct 649
    |||||
Db 686 ccgcaaggggacccacaaagagcttcttagaagaggacacccgctgcaggccatccggtgacct 745

Qy 650 gctgtccccccgaaagaccacttactgtatcatctcagggccctctcggcagctgtaggg 709
    |||||
Db 746 gctgtccccccgaaagaccacttactgtatcatctcagggccctctcggcagctgtaggg 805

Qy 710 gtggggaccggggagacctgctgtagcccccatcagacctgcccaagcaccatag 769
    |||||
Db 806 gtggggaccggggagacctgctgtagcccccatcagacctgcccaagcaccatag 865

Qy 770 gaaataaagtcttttc 785
    |||||
Db 866 gaaataaagtcttttc 881
    |||||
```

Wed May 24 09:19:46 2000

|||||
Db 686 ccgcaaggaccacaaagacgctctagaaaggggaccgagagctctccactccag 745
QY 650 gctgtcccccgaagaccacttactgtacatctctcagccctctcggcagctaggg 709
Db 746 gctgtcccccgaagaccacttactgtacatctctcagccctctcggcagctaggg 805
QY 710 gtggggaccgggagacactgtctgtagcccccacagaccctgcccacgaccatag 769
Db 806 gtggggaccgggagacactgtctgtagcccccacagaccctgcccacgaccatag 865
QY 770 gaaataaagtcttctc 785
Db 866 gaaataaagtcttctc 881

RESULT 12
PCT-US99-13418-118
; Sequence 118, Application PC/TUS99113418
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: 94 Human secreted proteins
; FILE REFERENCE: P2029PCT
; CURRENT APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER FILING DATE: 1998-06-16
; EARLIER FILING DATE: 1998-06-16
; EARLIER FILING DATE: 1998-06-16
; EARLIER FILING DATE: 1998-06-16
; EARLIER FILING DATE: 1998-06-16
; EARLIER FILING DATE: 1998-06-22
; EARLIER FILING DATE: 1998-06-22
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 502
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 118
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (117)
; OTHER INFORMATION: n equals a,t,g, or c
PCT-US99-13418-118

Query Match 76.0%; Score 596.6; DB 1; Length 882;
Best Local Similarity 88.1%; Pred. No. 2.6e-117;
Matches 721; Conservative 1; Mismatches 1; Indels 95; Gaps 3;
QY 1 cgggaatcggagtcgggtgacccgggctgtgtctagcataaaaggcgagccccaaga 60
Db 67 cgggaatcggagtcgggtgacccgggctgtgtctagcataaaaggcgga-nccagaaga 125
QY 61 agggcggggtatgggagagcctcccaactgcccccgcaaggcggaatcgtggtcc 120
Db 126 agggcggggtatgggagagcctcccaactgcccccgcaaggcggaatcgtggtcc 185
QY 121 tgcgtgctctctctacactgtgtatccctccgctgagctctatccatgatgctg 180
Db 186 tgcgtgctctctctacactgtgtatccctccgctgagctctatccatgatgctg 245
QY 181 acgccaagagagctccttgggttcacagcctccagcctcactcactcaggttcagcc 240
Db 246 acgccaagagagctccttgggttcacagcctccagcctcactcactcaggttcagcc 305
QY 241 gacttttctgaaaggttaacctgtctcgggcatagacagcttattctgtgccccatgg 300
Db 306 gacttttctgaaa-gtaacctgtctcgggcatagacagcttattctgtgccccatgg 364

QY 301 acttcgggggctccctgggaaactaccacaaagagagaccagagaccagctggggga 360
Db 365 acttcgggggctccctgggaaactaccacaaagagagaccagagaccagctggggga 424
QY 361 acaacacctctcagccacctccagatgcac----- 393
Db 425 acaacacctctcagccacctccagatgcacaaagatgaccgacaacagagagaggg 484
QY 393 ----- 393
Db 485 tgcgtatctccgagaatgtgtggcatctcaacacagcgagggttcgaggggtg 544
QY 393 -----aaggtaccagagatggagagagagggccctgggtaccctccagaaggccacgg 447
Db 545 atttgaagggtaccacagagatggagagagagggccctgggtaccctccagaaggccacgg 604
QY 448 acagcttcacacagaaactccatcccccgggtgcccctctctggatcatttaagctccacggc 507
Db 605 acagcttcacacagaaactccatcccccgggtgcccctctctggatcatttaagctccacggc 664
QY 508 ggaggtccacacagagatgctgagggcgccactggtcagcgagagagagagagagggga 567
Db 665 ggaggtccacacagagatgctgagggcgccactggtcagcgagagagagagagagggga 724
QY 568 tgcagggccatccgggagtgactccgcaagggggacccacaaaggagctctctagaagagggga 627
Db 725 tgcagggccatccgggagtgactccgcaagggggacccacaaaggagctctctagaagagggga 784
QY 628 ccgagagctctcccaactccaggtgtcccccgcgaagagaccacttactgtacatctca 687
Db 785 ccgagagctctcccaactccaggtgtcccccgcgaagagaccacttactgtacatctca 844
QY 688 gcccctctcggcagctgtagggtggggaccggggagc 725
Db 845 gcccctctcggcagctgtagggtggggaccggggagc 882

RESULT 13

US-09-009-325-15
; Sequence 15, Application US/09009325A
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-008DV3
; CURRENT APPLICATION NUMBER: US/09/009,325A
; EARLIER FILING DATE: 1998-01-20
; EARLIER FILING DATE: 1997-04-17
; EARLIER FILING DATE: 1998-01-15
; EARLIER FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(726)
US-09-009-325-15

Query Match 67.5%; Score 530; DB 35; Length 726;
Best Local Similarity 87.2%; Pred. No. 3.5e-103;
Matches 633; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 72 atgggagaagctctcccaactgcccccgcaaggcggaatctgtgtctctgtcgtcgtc 131
Db 1 atgggagaagctctcccaactgcccccgcaaggcggaatctgtgtcgtcgtcgtc 60
QY 132 ctctctacacctgtgtatccctccgctgagctctctatccatgatgtgacgccaagag 191

Db 61 ctctctacccctgggtgacccctccgctcagctcctatccatgatgctgagcccaagag 120
Qy 192 agctccttggtgtctcaagagcctccagagcctcactccaaggtctcagcgactttctcg 251
Db 121 agctccttggtgtctcaagagcctccagagcctcactccaaggtctcagcgactttctcg 180
Qy 252 aaaggttaacctgcttcgggcatagacagcttattctctgccccatgagcttcggggc 311
Db 181 aaaggttaacctgcttcgggcatagacagcttattctctgccccatgagcttcggggc 240
Qy 312 ctccctgggaactaccacaagagagagaccagagcagcagctgggggaacaacaccctc 371
Db 241 ctccctgggaactaccacaagagagagaccagagcagcagctgggggaacaacaccctc 300
Qy 372 tccagccacctccagatcgac----- 393
Db 301 tccagccacctccagatcgacaaagatgaccgacacaagacagagaggtgctgatctcc 360
Qy 393 -----aaggtta 398
Db 361 gagaatgtggtggcatccattcaaccagcgaggggagcttcgaggggtgattgaaggtta 420
Qy 399 ccaggtatggaggaagagggccctggtaccatccagaagggccacgagcagcttccac 458
Db 421 ccaggtatggaggaagagggccctggtaccatccagaagggccacgagcagcttccac 480
Qy 459 acagaactccatcccccgggtggtctcttgatcattaaagctgcacggcgaggttccac 518
Db 481 acagaactccatcccccgggtggtctcttgatcattaaagctgcacggcgaggttccac 540
Qy 519 cagatgctctggagggcgccactggctcagcagagagcagacacgcctcagggccatc 578
Db 541 caggtatgctctggagggcgccactggctcagcagagagcagacacgcctcagggccatc 600
Qy 579 cgggtatgactccgcaaggggaccacaaagagcgtccttagaagggggaccgagagctcc 638
Db 601 cgggtatgactccgcaaggggaccacaaagagcgtccttagaagggggaccgagagctcc 660
Qy 639 tccactcaggtctgccccccgaaagaccacttactgtacatctcagggccctctcgg 698
Db 661 tccactcaggtctgccccccgaaagaccacttactgtacatctcagggccctctcgg 720
Qy 699 cagctg 704
Db 721 cagctg 726

RESULT 14
US-09-685-15
; Sequence 15, Application US/09009685A
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE OF INVENTION: THEREFOR
; FILE REFERENCE: MEI-008DV2
; CURRENT APPLICATION NUMBER: US/09/009,685A
; CURRENT FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/843,704
; EARLIER FILING DATE: 1997-04-16
; EARLIER APPLICATION NUMBER: 08/842,898
; EARLIER FILING DATE: 1997-04-17
; EARLIER APPLICATION NUMBER: 60/071,589
; EARLIER FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)

US-09-009-685-15
Query Match 67.5%; Score 530; DB 35; Length 726;
Best Local Similarity 87.2%; Pred. No. 3.5e-103;
Matches 633; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
Qy 72 atgggagaagcctcccccacctgcccccaagcgcgacatctgctgctcgtcgtcgc 131
Db 1 atgggagaagcctcccccacctgcccccaagcgcgacatctgctgctcgtcgtcgc 60
Qy 132 ctctctacccctgggtgacccctccgctcagctcctatccatgatgctgagcccaagag 191
Db 61 ctctctacccctgggtgacccctccgctcagctcctatccatgatgctgagcccaagag 120
Qy 192 agctccttggtgtctcaagggcctccagagcctcactccaaggttcagcgactttctcg 251
Db 121 agctccttggtgtctcaagggcctccagagcctcactccaaggttcagcgactttctcg 180
Qy 252 aaaggttaacctgcttcgggcatagacagcttattctctgccccatgagcttcggggc 311
Db 181 aaaggttaacctgcttcgggcatagacagcttattctctgccccatgagcttcggggc 240
Qy 312 ctccctgggaactaccacaagagagagaccagagcagcagctgggggaacaacaccctc 371
Db 241 ctccctgggaactaccacaagagagagaccagagcagcagctgggggaacaacaccctc 300
Qy 372 tccagccacctccagatcgac----- 393
Db 301 tccagccacctccagatcgacaaagatgaccgacacaagacagagaggtgctgatctcc 360
Qy 393 -----aaggtta 398
Db 361 gagaatgtggtggcatccattcaaccagcgaggggagcttcgaggggtgattgaaggtta 420
Qy 399 ccaggtatggaggaagagggccctggtaccatccagaagggccacgagcagcttccac 458
Db 421 ccaggtatggaggaagagggccctggtaccatccagaagggccacgagcagcttccac 480
Qy 459 acagaactccatcccccgggtggtctcttgatcattaaagctgcacggcgaggttccac 518
Db 481 acagaactccatcccccgggtggtctcttgatcattaaagctgcacggcgaggttccac 540
Qy 519 cagatgctctggagggcgccactggctcagcagagagcagacacgcctcagggccatc 578
Db 541 caggtatgctctggagggcgccactggctcagcagagagcagacacgcctcagggccatc 600
Qy 579 cgggtatgactccgcaaggggaccacaaagagcgtccttagaagggggaccgagagctcc 638
Db 601 cgggtatgactccgcaaggggaccacaaagagcgtccttagaagggggaccgagagctcc 660
Qy 639 tccactcaggtctgccccccgaaagaccacttactgtacatctcagggccctctcgg 698
Db 661 tccactcaggtctgccccccgaaagaccacttactgtacatctcagggccctctcgg 720
Qy 699 cagctg 704
Db 721 cagctg 726

RESULT 15
US-09-009-706-15
; Sequence 15, Application US/09009706A
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE OF INVENTION: THEREFOR
; FILE REFERENCE: MEI-008DV1
; CURRENT APPLICATION NUMBER: US/09/009,706A
; CURRENT FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/842,898
; EARLIER FILING DATE: 1997-04-17
; EARLIER APPLICATION NUMBER: 60/071,589

Search completed: May 24, 2000, 01:02:32
Job time: 14324 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:42:13 ; Search time 6115.43 Seconds
(without alignments)
-135.688 Million cell updates/sec

Title: US-09-215-435-123
Perfect score: 853
Sequence: 1 ggaggatggcgagcagtct.....tttgcgaaaaaaaaaaaaa 853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*
51: gb_pr5.*
52: gb_htg8.*
53: gb_htg9.*
54: gb_htg10.*
55: gb_htg11.*
56: gb_htg12.*
57: gb_htg13.*
58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	568.2	66.6	1998	11	HSM800599
2	566.6	66.4	1592	51	AF161522
3	268.8	31.5	138262	57	AC021660
4	131.4	15.4	138262	57	AC021660
5	112.4	13.2	179245	40	HUAC002492
6	103	12.1	297663	44	AC008785
7	98.4	11.5	132041	41	AC010762
8	97.4	11.4	221184	57	AC012520
9	95.6	11.2	183982	32	AP001003
10	93.8	11.0	214736	32	AP001156
11	79.4	9.3	129057	44	AC016336
12	77.8	9.1	127000	33	AL138903
13	76.2	8.9	171671	54	AC011497
14	75.8	8.9	126352	10	HS23K20
15	75.6	8.9	220610	44	AC010404
16	75.4	8.8	137494	40	AC008127
17	75.4	8.8	304755	45	AC022025
18	75	8.8	137678	10	HS879K22
19	74.4	8.7	169724	52	AC021450
20	74	8.7	160626	40	AC006374
21	72.6	8.5	158602	33	AL133283
22	71.4	8.4	127933	45	AC008849
23	70.6	8.3	168977	45	AC022065
24	69.4	8.1	213455	32	AL135901
25	69.2	8.1	102311	40	AC005377
26	69.2	8.1	194133	41	AC009502
27	67.8	7.9	132155	32	HS3300G12
28	67.4	7.9	163193	55	AC022562
29	67.2	7.9	153741	55	AC011789
30	66.6	7.8	161905	42	AC011034
31	66.2	7.8	160748	40	AC001234
32	66.2	7.8	165843	32	AP000772
33	66.2	7.8	201239	40	AF060568
34	66.2	7.8	204134	32	AP000774
35	66	7.7	171455	41	AC007394
36	65.4	7.7	129057	44	AC016336
37	65	7.6	117392	11	CNS01DSM
38	64.8	7.6	175175	52	AC021538
39	64.8	7.6	191068	43	AC016913
40	64.4	7.5	122961	10	HSJ193N13
41	64.2	7.5	202550	33	AL139340
42	63.4	7.4	72321	43	AC016536
43	63.4	7.4	174890	52	AC020918
44	63.4	7.4	189120	44	AC008595
45	63	7.4	105887	45	AC022102

ALIGNMENTS

Wed May 24 09:19:47 2000

us-09-215-435-123.rge

```

|||||
Db 131 TTCTGGTATGAATATCGAAGTCCAGTTCAGAAAATTCAGTATTTGAAATAAAGCATC 190
|||||
Qy 181 tgggatgaattcattagtagatgaagcagatgaagaaccttataatgatgcacattttcga 240
|||||
Db 191 TGGGATGAATTCATTAGTATGAGGAGATCAAAAGACTTTATAATGATGACATTTTCGA 250
|||||
Qy 241 tacaatggcacagtggtgattgtgagagcgtgtatcaccatcccaaaaacatgcattgg 300
|||||
Db 251 TACAATGGCACAGTGGGATTGTGGAGACGGTGTATCACCATACCCAAAACATGCATTGG 310
|||||
Qy 301 tatagccaccagaaagacagagtcatttggatggtgtcacaataatgtgtgattttcaca 360
|||||
Db 311 TATAGCCACAGAAAGACAGAGTCATTTGATGTTGGTGCACAAAATGTGAGTTTCACA 370
|||||
Qy 361 ctaactgagcagtcattcagagaaatttggatgcccccagaaacacacacagcagattgat 420
|||||
Db 371 CTAAGTGGACAGTTCATCGAGAAATTTGTTGATCCCGAAACCAATAGCGGATTGAT 430
|||||
Qy 421 ctcttaggacctatcttggcgttgcagttccttcccttttacccttttggattgagtttg 480
|||||
Db 431 CTCTTAGACACTATCTTGGCGTGGCAGTTCCTTTACCTTTTGTGAGTTTAGGTTTG 490
|||||
Qy 481 atgtgcttggggcttggatcggaacttggctgtcgttcattgcccgaagcctatatccacc 540
|||||
Db 491 ATGTGCTTTGGGCTTTGATCGGACTTTGTGCTTGCATTTGCCGAAGCTTATATCCACC 550
|||||
Qy 541 attgccacggcattctcctcctccttccttcagat 573
|||||
Db 551 ATTGCCACGGGCATTCTCCATCTCCTTGCAGGT 583
|||||

RESULT 2
AF161522 1592 bp mRNA PRI 01-FEB-2000
LOCUS Homo sapiens HSPC174 mRNA, complete cds.
DEFINITION AF161522
ACCESSION AF161522
VERSION AF161522.1 GI:6841567
KEYWORDS FLI CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1592)
AUTHORS Ye,M., Zhang,Q.H., Zhou,J., Shen,Y., Wu,X.Y., Guan,Z.Q., Wang,L.,
Fan,H.Y., Mao,Y.F., Dai,M., Huang,Q.H., Chen,S.J. and Chen,Z.
Human full length cDNA cloned from cd34+ stem cells
Unpublished
TITLE 2 (bases 1 to 1592)
JOURNAL Ye,M., Zhang,Q.H., Zhou,J., Shen,Y., Wu,X.Y., Guan,Z.Q., Wang,L.,
Fan,H.Y., Mao,Y.F., Dai,M., Huang,Q.H., Chen,S.J. and Chen,Z.
Direct Submission
SUBMITTED (23-JUN-1999) Shanghai Institute of Hematology, Shanghai
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P. R. China
FEATURES
Location/Qualifiers
1..1592
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CBNAPD08"
/cell_type="cd34+ stem cells"
/tissue_type="blood"
131..826
/codon_start=1
/product="HSPC174"
/protein_id="AAF29137.1"
/db_xref="GI:6841568"
/translaton="MAASIGTDFWYERSPVOENSDLNKSIWDEFISDEADEKTYND
ALFNGVGTGLMRRRCITIPKNNHWYSPPERTESFDVTKVFSFTLEQFMKFDVPGN
HNSGIDLLTYLWRCOFLPLPFYSLGCMGALIGLCACICRSLYPTIATGILHLAAGL
CTIGSVSCYVAGIELLHQKLELFDNVSGEFWSFCLACVSAPLQFMASALFIWAHTN
RKREYTLMKAYRVA"
BASE COUNT 447 a 307 c 320 g 518 t

Qy 1 gagagatggcgagcagctgaatccagaatggataacccttttgcacagcatttga 60
|||||
Db 11 GGAGGATGGCGAGCAGTCTGAATCCAGATGGATAACCGTTTGGCTACAGCATTTGTA 70
|||||
Qy 61 attctgtgtgcttagcctcattccaccatctacatggcagcctccattggcacagac 120
|||||
Db 71 ATGTGTTGTGCTTAGCTCTATTTCCACCATCTACATGGCAGCCTCCATTGGCAGAC 130
|||||
Qy 121 ttctggtgaatcgaaagtcagtcacgaataatccagaaattccagtgattgtaataaagc 180
|||||

```


* 13958	13977: gap of unknown length	* 64071	66627: contig of 2557 bp in length
* 13978	14528: contig of 551 bp in length	* 66628	66647: gap of unknown length
* 14529	14548: gap of unknown length	* 66648	68423: contig of 1776 bp in length
* 14549	14715: contig of 167 bp in length	* 68443	68443: gap of unknown length
* 14716	14735: gap of unknown length	* 71116	71116: contig of 2673 bp in length
* 14736	15507: contig of 772 bp in length	* 71137	71136: gap of unknown length
* 15508	15527: gap of unknown length	* 73020	73039: contig of 1883 bp in length
* 15528	16561: contig of 1034 bp in length	* 73040	73039: gap of unknown length
* 16562	16581: gap of unknown length	* 74951	74951: contig of 1912 bp in length
* 16582	17696: contig of 1115 bp in length	* 74952	74971: gap of unknown length
* 17697	17716: gap of unknown length	* 74972	76581: contig of 1610 bp in length
* 17717	19280: contig of 1564 bp in length	* 76582	76601: gap of unknown length
* 19281	19300: gap of unknown length	* 76602	78478: contig of 1877 bp in length
* 19301	20370: contig of 1070 bp in length	* 78479	78498: gap of unknown length
* 20371	20390: gap of unknown length	* 81741	81740: contig of 3242 bp in length
* 20391	21168: contig of 778 bp in length	* 81741	81760: gap of unknown length
* 21169	22188: gap of unknown length	* 84837	84837: contig of 3077 bp in length
* 21189	22491: contig of 1303 bp in length	* 84838	84857: gap of unknown length
* 22492	22511: gap of unknown length	* 86992	86992: contig of 2135 bp in length
* 22512	23450: contig of 939 bp in length	* 86993	87012: gap of unknown length
* 23451	23470: gap of unknown length	* 87013	90043: contig of 3031 bp in length
* 23471	24810: contig of 1340 bp in length	* 90044	90063: gap of unknown length
* 24811	24830: gap of unknown length	* 90064	92958: contig of 2895 bp in length
* 24831	26823: contig of 1993 bp in length	* 92959	92978: gap of unknown length
* 26824	26843: gap of unknown length	* 96579	96579: contig of 3601 bp in length
* 26844	28212: contig of 1369 bp in length	* 96580	96599: gap of unknown length
* 28213	28232: gap of unknown length	* 96600	100688: contig of 4089 bp in length
* 28233	29157: contig of 925 bp in length	* 100689	100708: gap of unknown length
* 29158	29177: gap of unknown length	* 100709	104507: contig of 3799 bp in length
* 29178	30368: contig of 1191 bp in length	* 104508	104527: gap of unknown length
* 30369	30388: gap of unknown length	* 104528	107058: contig of 2531 bp in length
* 30389	31656: contig of 1268 bp in length		
* 31657	31676: gap of unknown length		
* 31677	32803: contig of 1127 bp in length		
* 32804	32823: gap of unknown length		
* 32824	33900: contig of 1067 bp in length		
* 33891	33910: gap of unknown length		
* 33911	35184: contig of 1274 bp in length		
* 35185	35204: gap of unknown length		
* 35205	36891: contig of 1687 bp in length		
* 36892	36911: gap of unknown length		
* 36912	38109: contig of 1198 bp in length		
* 38110	38129: gap of unknown length		
* 38130	39345: contig of 1216 bp in length		
* 39346	39365: gap of unknown length		
* 39366	41036: contig of 1671 bp in length		
* 41037	41056: gap of unknown length		
* 41057	42331: contig of 1275 bp in length		
* 42332	42351: gap of unknown length		
* 42352	44275: contig of 1924 bp in length		
* 44276	44295: gap of unknown length		
* 44296	45668: contig of 1273 bp in length		
* 45669	45689: gap of unknown length		
* 45689	47071: contig of 1483 bp in length		
* 47072	47091: gap of unknown length		
* 47092	48632: contig of 1541 bp in length		
* 48633	48652: gap of unknown length		
* 48653	50719: contig of 2067 bp in length		
* 50720	50739: gap of unknown length		
* 50740	52382: contig of 1643 bp in length		
* 52383	52402: gap of unknown length		
* 52403	53708: contig of 1306 bp in length		
* 53709	53728: gap of unknown length		
* 53729	56249: contig of 2521 bp in length		
* 56250	56269: gap of unknown length		
* 56270	57988: contig of 1719 bp in length		
* 57989	58008: gap of unknown length		
* 58009	60092: contig of 2084 bp in length		
* 60093	60112: gap of unknown length		
* 60113	61863: contig of 1751 bp in length		
* 61864	61883: gap of unknown length		
* 64051	64050: contig of 2167 bp in length		
	64070: gap of unknown length		
<hr/>			
Query Match 31.5%; Score 268.8; DB 57; Length 138262;			
Best Local Similarity 95.8%; Pred. No. 3.1e-64;			
Matches 276; Conservative 0; Mismatches 12; Indels 0; Gaps 0;			
QY 552	cattctcatctccttgagatataccatctgtgaagtcagccacatggaggtcctg 611		
DB 59884	CATTCCCTCTCCAAACCCGATACCATCTGTGAAGTCCAGCCACATGGAGGTCTCTG 59825		
QY 612	tgtatagctccagctgaatcccaagcttaagctcccaactgacagccacacatcttcc 671		
DB 59824	TGTATAGCTCCAGCTGAATCCCAAGCTAAGCTCCCAACTGACAGCCCAACATCATTTCC 59765		
QY 672	agccatgtgtggagccatcctctggatgtccagccttaacagccttcagagactcagc 731		
DB 59764	AGCCATGTGTGGAGCCATCTCTGGATGTCCAGCCTTAACAAGCCTTCAGATGACTTCAGC 59705		
QY 732	cacagctattacttactacatcctgtgagacttaataaagaaccaactagctgagcc 791		
DB 59704	CACAGCTATTACTTACTACATCCTTGTGAGACTCTATATAAGAACCACCACTAGCTGAGCC 59645		
QY 792	caatcaacctatggaactgatagaaataaataaattgtgtgttgcg 839		
DB 59644	CAATCAACCTATGGAACCTGGTAGAAATAAATAAATGAATTGTTGTTTGTG 59597		
RESULT 4			
AC021660	138262 bp DNA HTG 19-FEB-2000		
LOCUS	Homo sapiens clone RP11-227H4, *** SEQUENCING IN PROGRESS ***		
DEFINITION	unordered pieces.		
ACCESSION	AC021660		
VERSION	AC021660.4 GI:7007887		
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
	1 (bases 1 to 138262)		
	Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,		
	Bodoty,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,		
	Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,		

David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L., Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hognes, M., Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, W., Logan, O., Lozado, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswald, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R., Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 138262)
Worley, K.C.

Direct Submission
Submitted (19-JAN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Feb 19, 2000 this sequence version replaced gi:6751635.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMUR
Center clone name: RP11-227H4
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 81131 bases at least Q40
Consensus quality: 99645 bases at least Q30
Consensus quality: 109526 bases at least Q20
Estimated insert size: 121480; sum-of-contigs estimation
Quality coverage: 1.3x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 811: contig of 811 bp in length
* 812: gap of unknown length
* 832: contig of 771 bp in length
* 1603: gap of unknown length
* 1623: contig of 775 bp in length
* 2398: gap of unknown length
* 2418: contig of 1220 bp in length
* 3637: gap of unknown length
* 3658: contig of 777 bp in length
* 4434: gap of unknown length
* 4435: contig of 768 bp in length
* 5222: gap of unknown length
* 5223: contig of 813 bp in length
* 5243: gap of unknown length
* 6075: contig of 818 bp in length
* 6076: gap of unknown length
* 6894: contig of 795 bp in length
* 6914: gap of unknown length
* 7709: contig of 786 bp in length
* 7729: gap of unknown length
* 8514: contig of 1322 bp in length
* 8535: gap of unknown length
* 9857: gap of unknown length

10629: contig of 753 bp in length
10649: gap of unknown length
11446: contig of 797 bp in length
11466: gap of unknown length
12349: contig of 883 bp in length
12369: gap of unknown length
12370: contig of 858 bp in length
12378: gap of unknown length
13247: contig of 710 bp in length
13957: gap of unknown length
13977: contig of 551 bp in length
14528: gap of unknown length
14548: contig of 167 bp in length
14715: gap of unknown length
14735: contig of 772 bp in length
15507: gap of unknown length
15527: contig of 1034 bp in length
15628: gap of unknown length
16581: contig of 1115 bp in length
17696: gap of unknown length
17716: contig of 1564 bp in length
19280: gap of unknown length
19300: contig of 1070 bp in length
20370: gap of unknown length
20390: contig of 778 bp in length
21168: gap of unknown length
21188: contig of 1303 bp in length
22491: gap of unknown length
22511: contig of 939 bp in length
23450: gap of unknown length
23470: contig of 1340 bp in length
24810: gap of unknown length
24830: contig of 1993 bp in length
26823: gap of unknown length
26843: contig of 1369 bp in length
28212: gap of unknown length
28232: contig of 925 bp in length
29157: gap of unknown length
29177: contig of 1191 bp in length
30368: gap of unknown length
30388: contig of 1268 bp in length
31656: gap of unknown length
31676: contig of 1127 bp in length
32803: gap of unknown length
32823: contig of 1067 bp in length
33910: gap of unknown length
33918: contig of 1274 bp in length
35204: gap of unknown length
35205: contig of 1687 bp in length
36891: gap of unknown length
36911: contig of 1198 bp in length
38109: gap of unknown length
38129: contig of 1216 bp in length
39345: gap of unknown length
39365: contig of 1671 bp in length
41036: gap of unknown length
41056: contig of 1275 bp in length
42331: gap of unknown length
42351: contig of 1924 bp in length
44275: gap of unknown length
44295: contig of 1273 bp in length
45668: gap of unknown length
45688: contig of 1483 bp in length
47071: gap of unknown length
47091: contig of 1541 bp in length
48632: gap of unknown length
48652: contig of 2067 bp in length
50719: gap of unknown length
50739: contig of 1643 bp in length
52382: gap of unknown length
52402: contig of 1306 bp in length
53708: gap of unknown length
53728: contig of 5521 bp in length
56249: gap of unknown length

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

RESULT 6

AC008785
LOCUS AC008785
DEFINITION Homo sapiens chromosome 16 clone C17B-H1_2036A2, WORKING DRAFT
SEQUENCE, 136 unordered pieces.
ACCESSION AC008785
VERSION
KEYWORDS
SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 297663)
DOE Joint Genome Institute.
TITLE
JOURNAL
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 14, 2000 this sequence version replaced gi:5686286.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>
-----Summary Statistics
Consensus quality: 157653 bases at least Q40
Consensus quality: 218991 bases at least Q30
Consensus quality: 237343 bases at least Q20
Estimated insert size: 297663; sum-of-contigs estimation
Estimated insert size: 164000; agarose-fp estimation
Quality coverage: 4.02x in Q20 bases; agarose-fp estimation
Quality coverage: 2.21x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 136 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1171: contig of 1171 bp in length
* gap of unknown length
* 1172 2220: contig of 1049 bp in length
* gap of unknown length
* 2221 3231: contig of 1011 bp in length
* gap of unknown length
* 3232 4235: contig of 1004 bp in length
* gap of unknown length
* 4236 5262: contig of 1027 bp in length
* gap of unknown length
* 5263 6457: contig of 1195 bp in length
* gap of unknown length
* 6458 7847: contig of 1390 bp in length
* gap of unknown length
* 7848 8913: contig of 1066 bp in length
* gap of unknown length
* 8914 10901: contig of 1988 bp in length
* gap of unknown length
* 10902 12278: contig of 1377 bp in length
* gap of unknown length
* 12279 13806: contig of 1528 bp in length
* gap of unknown length
* 13807 15510: contig of 1704 bp in length
* gap of unknown length
* 15511 17069: contig of 1559 bp in length
* gap of unknown length
* 17070 18220: contig of 1151 bp in length
* gap of unknown length
*
* 18221 19329: contig of 1109 bp in length
* gap of unknown length
* 19330 20609: contig of 1280 bp in length
* gap of unknown length
* 20610 21663: contig of 1054 bp in length
* gap of unknown length
* 21664 22864: contig of 1201 bp in length
* gap of unknown length
* 22865 24284: contig of 1420 bp in length
* gap of unknown length
* 24285 25400: contig of 1116 bp in length
* gap of unknown length
* 25401 27665: contig of 2265 bp in length
* gap of unknown length
* 27666 28768: contig of 1103 bp in length
* gap of unknown length
* 28769 30006: contig of 1238 bp in length
* gap of unknown length
* 30007 31209: contig of 1203 bp in length
* gap of unknown length
* 31210 32381: contig of 1172 bp in length
* gap of unknown length
* 32382 33537: contig of 1156 bp in length
* gap of unknown length
* 33538 34606: contig of 1069 bp in length
* gap of unknown length
* 34607 35693: contig of 1087 bp in length
* gap of unknown length
* 35694 36761: contig of 1068 bp in length
* gap of unknown length
* 36762 38267: contig of 1506 bp in length
* gap of unknown length
* 38268 39319: contig of 1052 bp in length
* gap of unknown length
* 39320 40442: contig of 1123 bp in length
* gap of unknown length
* 40443 41926: contig of 1484 bp in length
* gap of unknown length
* 41927 43031: contig of 1105 bp in length
* gap of unknown length
* 43032 44239: contig of 1208 bp in length
* gap of unknown length
* 44240 45770: contig of 1531 bp in length
* gap of unknown length
* 45771 47203: contig of 1433 bp in length
* gap of unknown length
* 47204 48254: contig of 1051 bp in length
* gap of unknown length
* 48255 49771: contig of 1517 bp in length
* gap of unknown length
* 49772 50837: contig of 1066 bp in length
* gap of unknown length
* 50838 52277: contig of 1440 bp in length
* gap of unknown length
* 52278 53931: contig of 1654 bp in length
* gap of unknown length
* 53932 55000: contig of 1089 bp in length
* gap of unknown length
* 55001 56174: contig of 1174 bp in length
* gap of unknown length
* 56175 57431: contig of 1257 bp in length
* gap of unknown length
* 57432 58510: contig of 1079 bp in length
* gap of unknown length
* 58511 60377: contig of 1867 bp in length
* gap of unknown length
* 60378 61795: contig of 1418 bp in length
* gap of unknown length
* 61796 62807: contig of 1012 bp in length
* gap of unknown length
* 62808 63892: contig of 1085 bp in length
* gap of unknown length
* 63893 64902: contig of 1010 bp in length
* gap of unknown length

* 64903	gap of 1369 bp in length	66271:	contig of 1369 bp in length
* 66272	gap of unknown length	68214:	contig of 1943 bp in length
* 68215	gap of unknown length	70612:	contig of 2398 bp in length
* 70613	gap of unknown length	71673:	contig of 1061 bp in length
* 71674	gap of unknown length	73301:	contig of 1628 bp in length
* 73302	gap of unknown length	74709:	contig of 1408 bp in length
* 74710	gap of unknown length	76210:	contig of 1501 bp in length
* 76211	gap of unknown length	77529:	contig of 1319 bp in length
* 77530	gap of unknown length	79208:	contig of 1679 bp in length
* 79209	gap of unknown length	81534:	contig of 2326 bp in length
* 81535	gap of unknown length	82606:	contig of 1072 bp in length
* 82607	gap of unknown length	84386:	contig of 1780 bp in length
* 84387	gap of unknown length	86136:	contig of 1750 bp in length
* 86137	gap of unknown length	88065:	contig of 1929 bp in length
* 88066	gap of unknown length	90535:	contig of 2470 bp in length
* 90536	gap of unknown length	91995:	contig of 1460 bp in length
* 91996	gap of unknown length	93002:	contig of 1007 bp in length
* 93003	gap of unknown length	94777:	contig of 1775 bp in length
* 94778	contig of 2380 bp in length	97157:	contig of 2380 bp in length
* 97158	gap of unknown length	98340:	contig of 1183 bp in length
* 98341	gap of unknown length	99365:	contig of 1025 bp in length
* 99366	gap of unknown length	101874:	contig of 2509 bp in length
* 101875	gap of unknown length	102911:	contig of 1037 bp in length
* 102912	gap of unknown length	104424:	contig of 1513 bp in length
* 104425	gap of unknown length	106452:	contig of 2028 bp in length
* 106453	contig of 2170 bp in length	108622:	contig of 2170 bp in length
* 108623	gap of unknown length	109639:	contig of 1017 bp in length
* 109640	gap of unknown length	110844:	contig of 1205 bp in length
* 110845	contig of 1018 bp in length	111862:	contig of 1018 bp in length
Query Match 12.1%; Score 103; DB 44; Length 297663;			
Best Local Similarity 71.1%; Pred. No. 5.le-18;			
Matches 150; Conservative 0; Mismatches 60; Indels 1; Gaps 1;			
Qy 582	gtgaagtcacaggccacatgagtgctcctgtgtagctgtccagctggaatcccaagcta 641		
Db 53577	GAGAAGCCCAACACATAGAAAGCCACGTGAAGTGTCTCAGGCGACAGCCCAACTG 53636		
Qy 642	agctcccaactgacagccacatcattccagccatgtgtgaggagccatcctggatgtcc 701		
Db 53637	AGCTCCTAACCAAGTAGCCATATCATG-CCAGCCATGTGTCTCAGCCATCTTGGATGCC 53695		
Qy 702	agccttaaacagcttcagaggacttcagccacagctattattcttactacatcctgtga 761		

Db 53696	ACCCTAGTCAAGCCTTCAGATGACTCCAGCCCGAGTCATTATCTGACTACAACTGCATAG 53755		
Qy 762	gactctaataaagaaccactagctgagccc 792		
Db 53756	GACTTGAATGAGACACCGTAGGTGAAGCC 53786		
RESULT 7			
AC010762/c	132041 bp	DNA	HTG 22-SEP-1999
LOCUS	Homo sapiens chromosome 11 clone 46_D_05 map 11, *** SEQUENCING IN		
DEFINITION	PROGRESS ***, 29 unordered pieces.		
ACCESSION	AC010762.1	GI:5918000	
VERSION	HTG; HTGS_PHASE1.		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 132041)		
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Homo sapiens chromosome 11, clone 46_D_05		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 132041)		
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,		
	Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,		
	Brown, A., Castle, A., Collangelo, M., Collins, S., Collamore, A.,		
	Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,		
	Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,		
	Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,		
	Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,		
	Lehoczky, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,		
	McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,		
	Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,		
	Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,		
	Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,		
	Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,		
	Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.		
	Direct Submission		
TITLE	Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome		
JOURNAL	Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. &		
	Green, P. (1996-1997)		
	http://ftp.genome.washington.edu/RM/RepeatMasker.html.		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 29 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	1 2157: contig of 2157 bp in length		
	1 gap of unknown length		
	* 2158 4626: contig of 2469 bp in length		
	* gap of unknown length		
	* 4627 7180: contig of 2554 bp in length		
	* gap of unknown length		
	* 7181 11402: contig of 4222 bp in length		
	* gap of unknown length		
	* 11403 15093: contig of 3691 bp in length		
	* gap of unknown length		
	* 15094 17730: contig of 2637 bp in length		
	* gap of unknown length		
	* 17731 21289: contig of 3559 bp in length		
	* gap of unknown length		
	* 21290 24121: contig of 2832 bp in length		
	* gap of unknown length		
	* 24122 27013: contig of 2892 bp in length		
	* gap of unknown length		
	* 27014 32410: contig of 5397 bp in length		
	* gap of unknown length		
	* 32411 35020: contig of 2610 bp in length		

AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens 183,982 genomic DNA of 11q23
 Published only in DataBase (2000) In press
 2 (bases 1 to 183982)

REFERENCE

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (05-JAN-2000) to the DDBJ/EMBL/Genbank databases.
 Masahira Hattori, The Institute of Physical and Chemical Research
 (RIKEN), Genomic Sciences Center (GSC): Kitasato Univ., 1-15-1
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-42-778-9923, Fax:81-42-778-9924)
 On Feb 18, 2000 this sequence version replaced gi:6683676.

COMMENT

----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft11

Center clone name: RP11-680C5

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amerham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 149722 bases at least Q40
 Consensus quality: 161060 bases at least Q30
 Consensus quality: 167405 bases at least Q20
 Insert size: 170520; sum-of-contigs
 Quality coverage: 4.30x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of
 29 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

```

1 18031 contig of 18031 of in length
18532 34054 contig of 15523 of in length
34555 48195 contig of 13641 of in length
48696 59437 contig of 10742 of in length
59938 70966 contig of 11029 of in length
71467 81525 contig of 10059 of in length
82026 93042 contig of 11017 of in length
93543 101682 contig of 8140 of in length
102183 110472 contig of 8290 of in length
110973 118966 contig of 7994 of in length
119467 127714 contig of 8248 of in length
128215 133786 contig of 5572 of in length
134287 137340 contig of 3054 of in length
137841 142279 contig of 4439 of in length
142780 146346 contig of 3567 of in length
146847 150422 contig of 3576 of in length
150923 153700 contig of 2778 of in length
154201 157296 contig of 3096 of in length
154797 161257 contig of 3461 of in length
161758 164624 contig of 2867 of in length
165125 166950 contig of 1826 of in length
167451 169894 contig of 2444 of in length
170395 172335 contig of 1841 of in length
172736 174076 contig of 1341 of in length
174577 176002 contig of 1426 of in length
176503 178125 contig of 1623 of in length
178626 180174 contig of 1549 of in length
180675 182269 contig of 1595 of in length
182770 183982 contig of 1213 of in length.
```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 29 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 18029: contig of 18029 bp in length
18030 18531: gap of 502 bp
18532 34054: contig of 15523 bp in length
34055 34554: gap of 500 bp
34555 48192: contig of 13638 bp in length
48193 48696: gap of 504 bp
48697 59436: contig of 10740 bp in length
59437 59942: gap of 506 bp
59943 70965: contig of 11023 bp in length
70966 71470: gap of 505 bp
71471 81520: contig of 10050 bp in length
81521 82025: gap of 505 bp
82026 93039: contig of 11014 bp in length
93040 93547: gap of 508 bp
93548 101681: contig of 8134 bp in length
101682 102182: gap of 501 bp
102183 110471: contig of 8289 bp in length
110472 110984: gap of 513 bp
110985 118966: contig of 7982 bp in length
118967 119466: gap of 500 bp
119467 127713: contig of 8247 bp in length
127714 128215: gap of 502 bp
128216 133781: contig of 5566 bp in length
133782 134286: gap of 505 bp
134287 137340: contig of 3054 bp in length
137341 137841: gap of 501 bp
137842 142278: contig of 4437 bp in length
142279 142781: gap of 503 bp
142782 146346: contig of 3565 bp in length
146347 146848: gap of 502 bp
146849 150419: contig of 3571 bp in length
150420 150923: gap of 504 bp
150924 153700: contig of 2777 bp in length
153701 154201: gap of 501 bp
154202 157296: contig of 3095 bp in length
157297 157801: gap of 505 bp
157802 161248: contig of 3447 bp in length
161249 161764: gap of 516 bp
161765 164623: contig of 2859 bp in length
164624 165125: gap of 502 bp
165126 166950: contig of 1825 bp in length
166951 167453: gap of 503 bp
167454 169899: contig of 2436 bp in length
169899 170410: gap of 521 bp
170411 172233: contig of 1823 bp in length
172234 172735: gap of 502 bp
172736 174076: contig of 1341 bp in length
174077 174576: gap of 500 bp
174577 176001: contig of 1425 bp in length
176002 176502: gap of 501 bp
176503 178123: contig of 1621 bp in length
178124 178629: gap of 506 bp
178630 180174: contig of 1545 bp in length
180175 180674: gap of 500 bp
180675 182269: contig of 1595 bp in length
182270 182769: gap of 500 bp
182770 183982: contig of 1213 bp in length.
```

FEATURES
 source

```

1..183982
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-680C5"
/map="11q23"
```

BASE COUNT 44773 a 38911 c 40251 g 43470 t 16577 others
 ORIGIN

Query Match

11.2%; Score 95.6; DB 32; Length 183982;

NOTE: This is a 'working draft' sequence. It currently consists of 42 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as it becomes available and the accession number will be

as soon as the
preserved

Contig	Length (bp)
20627	11090
32217	11090

	contig	bp	length
32718	41231	8514	in length
32719	41232	8514	in length
32720	41233	8514	in length
32721	41234	8514	in length
32722	41235	8514	in length
32723	41236	8514	in length
32724	41237	8514	in length
32725	41238	8514	in length
32726	41239	8514	in length
32727	41240	8514	in length
32728	41241	8514	in length
32729	41242	8514	in length
32730	41243	8514	in length
32731	41244	8514	in length
32732	41245	8514	in length
32733	41246	8514	in length
32734	41247	8514	in length
32735	41248	8514	in length
32736	41249	8514	in length
32737	41250	8514	in length
32738	41251	8514	in length
32739	41252	8514	in length
32740	41253	8514	in length
32741	41254	8514	in length
32742	41255	8514	in length
32743	41256	8514	in length
32744	41257	8514	in length
32745	41258	8514	in length
32746	41259	8514	in length
32747	41260	8514	in length
32748	41261	8514	in length
32749	41262	8514	in length
32750	41263	8514	in length
32751	41264	8514	in length
32752	41265	8514	in length
32753	41266	8514	in length
32754	41267	8514	in length
32755	41268	8514	in length
32756	41269	8514	in length
32757	41270	8514	in length
32758	41271	8514	in length
32759	41272	8514	in length
32760	41273	8514	in length
32761	41274	8514	in length
32762	41275	8514	in length
32763	41276	8514	in length
32764	41277	8514	in length
32765	41278	8514	in length
32766	41279	8514	in length
32767	41280	8514	in length
32768	41281	8514	in length
32769	41282	8514	in length
32770	41283	8514	in length
32771	41284	8514	in length
32772	41285	8514	in length
32773	41286	8514	in length
32774	41287	8514	in length
32775	41288	8514	in length
32776	41289	8514	in length
32777	41290	8514	in length
32778	41291	8514	in length
32779	41292	8514	in length
32780	41293	8514	in length
32781	41294	8514	in length
32782	41295	8514	in length
32783	41296	8514	in length
32784	41297	8514	in length
32785	41298	8514	in length
32786	41299	8514	in length
32787	41300	8514	in length
32788	41301	8514	in length
32789	41302	8514	in length
32790	41303	8514	in length
32791	41304	8514	in length
32792	41305	8514	in length
32793	41306	8514	in length
32794	41307	8514	in length
32795	41308	8514	in length
32796	41309	8514	in length
32797	41310	8514	in length
32798	41311	8514	in length
32799	41312	8514	in length
32800	41313	8514	in length
32801	41314	8514	in length
32802	41315	8514	in length
32803	41316	8514	in length
32804	41317	8514	in length
32805	41318	8514	in length
32806	41319	8514	in length
32807	41320	8514	in length
32808	41321	8514	in length
32809	41322	8514	in length
32810	41323	8514	in length
32811	41324	8514	in length
32812	41325	8514	in length
32813	41326	8514	in length
32814	4132		

Contig	Length (bp)
41732	9612
51433	9612

	contig of	bp in length
61545	68791	7247
68800	76357	7066

Contig	Length (bp)
83794	6937
contig of	
76858	

Accession number	Contig	Length (bp)
84295	cont19 of 90307	5830
91008	cont19 of 96837	5830

Accession	Length (bp)	Contig of
97338	5351	102688
97339	5076	102688
97340	5076	102688
97341	5076	102688
97342	5076	102688
97343	5076	102688
97344	5076	102688
97345	5076	102688
97346	5076	102688
97347	5076	102688
97348	5076	102688
97349	5076	102688
97350	5076	102688
97351	5076	102688
97352	5076	102688
97353	5076	102688
97354	5076	102688
97355	5076	102688
97356	5076	102688
97357	5076	102688
97358	5076	102688
97359	5076	102688
97360	5076	102688
97361	5076	102688
97362	5076	102688
97363	5076	102688
97364	5076	102688
97365	5076	102688
97366	5076	102688
97367	5076	102688
97368	5076	102688
97369	5076	102688
97370	5076	102688
97371	5076	102688
97372	5076	102688
97373	5076	102688
97374	5076	102688
97375	5076	102688
97376	5076	102688
97377	5076	102688
97378	5076	102688
97379	5076	102688
97380	5076	102688
97381	5076	102688
97382	5076	102688
97383	5076	102688
97384	5076	102688
97385	5076	102688
97386	5076	102688
97387	5076	102688
97388	5076	102688
97389	5076	102688
97390	5076	102688
97391	5076	102688
97392	5076	102688
97393	5076	102688
97394	5076	102688
97395	5076	102688
97396	5076	102688
97397	5076	102688
97398	5076	102688
97399	5076	102688
97400	5076	102688

Accession number	Contig of	4876 bp in length
103183	103184 contig of	
109665	114540 contig of	

	contig of	bp	in length
115041	119314	4274	bp
115042	125106	5292	bp

Contig ID	Contig Length (bp)	Contig Description
125607	5425	Contig of 5425 bp in length
131031	5425	Contig of 5425 bp in length

Accession	Conc'n of	bp in length
131532	137138	4619
137639	142257	4619

142758	147766	contig of	5009 bp	in length
142759	147767	contig of	3342 bp	in length
142760	147768	contig of	3342 bp	in length

Contig	Length (bp)
155904	3796
152109	3796

	contig of	bp	in length
156405	160842	4438	in length
161343	164519	3177	bp

Accession	Contig	Length (bp)
165020	167901	2882

Contig	Length (bp)
1684402	2619
172483	2619
175602	2619
1772984	2619

Contig ID	Contig Length (bp)	Contig Type
176103	3630	contig of
179732	3097	contig of
182228	3630	contig of
182229	3097	contig of
182230	3630	contig of
182231	3097	contig of
182232	3630	contig of
182233	3097	contig of
182234	3630	contig of
182235	3097	contig of
182236	3630	contig of
182237	3097	contig of
182238	3630	contig of
182239	3097	contig of
182240	3630	contig of
182241	3097	contig of
182242	3630	contig of
182243	3097	contig of
182244	3630	contig of
182245	3097	contig of
182246	3630	contig of
182247	3097	contig of
182248	3630	contig of
182249	3097	contig of
182250	3630	contig of
182251	3097	contig of
182252	3630	contig of
182253	3097	contig of
182254	3630	contig of
182255	3097	contig of
182256	3630	contig of
182257	3097	contig of
182258	3630	contig of
182259	3097	contig of
182260	3630	contig of
182261	3097	contig of
182262	3630	contig of
182263	3097	contig of
182264	3630	contig of
182265	3097	contig of
182266	3630	contig of
182267	3097	contig of
182268	3630	contig of
182269	3097	contig of
182270	3630	contig of
182271	3097	contig of
182272	3630	contig of
182273	3097	contig of
182274	3630	contig of
182275	3097	contig of
182276	3630	contig of
182277	3097	contig of
182278	3630	contig of
182279	3097	contig of
182280	3630	contig of
182281	3097	contig of
182282	3630	contig of
182283	3097	contig of
182284	3630	contig of
182285	3097	contig of
182286	3630	contig of
182287	3097	contig of
182288	3630	contig of
182289	3097	contig of
182290	3630	contig of
182291	3097	contig of
182292	3630	contig of
182293	3097	contig of
182294	3630	contig of
182295	3097	contig of
182296	3630	contig of
182297	3097	contig of
182298	3630	contig of
182299	3097	contig of
182300	3630	contig of
182301	3097	contig of
182302	3630	contig of
182303	3097	contig of
182304	3630	contig of
182305	3097	contig of
182306	3630	contig of
182307	3097	contig of
182308	3630	contig of
182309	3097	contig of
182310	3630	contig of
182311	3097	contig of
182312	3630	contig of
182313	3097	contig of
182314	3630	contig of
182315	3097	contig of
182316	3630	contig of
182317	3097	contig of
182318	3630	contig of
182319	3097	contig of
182320	3630	contig of
182321	3097	contig of
182322	3630	contig of
182323	3097	contig of
182324	3630	contig of
182325	3097	contig of
182326	3630	contig of
182327	3097	contig of
182328	3630	contig of
182329	3097	contig of
182330	3630	contig of
182331	3097	contig of
182332		

Contig	Length (bp)
182830	3184
186013	3184

Accession number	Contig of	bp in length
186514	188993	2402
188993	contig of	1861
191356	contig of	1861
188993	contig of	1861

Accession	Contig	Length
191857	193485	1629 bp
191858	193485	1629 bp
191859	193485	1629 bp
191860	193485	1629 bp
191861	193485	1629 bp
191862	193485	1629 bp
191863	193485	1629 bp
191864	193485	1629 bp
191865	193485	1629 bp
191866	193485	1629 bp
191867	193485	1629 bp
191868	193485	1629 bp
191869	193485	1629 bp
191870	193485	1629 bp
191871	193485	1629 bp
191872	193485	1629 bp
191873	193485	1629 bp
191874	193485	1629 bp
191875	193485	1629 bp
191876	193485	1629 bp
191877	193485	1629 bp
191878	193485	1629 bp
191879	193485	1629 bp
191880	193485	1629 bp
191881	193485	1629 bp
191882	193485	1629 bp
191883	193485	1629 bp
191884	193485	1629 bp
191885	193485	1629 bp
191886	193485	1629 bp
191887	193485	1629 bp
191888	193485	1629 bp
191889	193485	1629 bp
191890	193485	1629 bp
191891	193485	1629 bp
191892	193485	1629 bp
191893	193485	1629 bp
191894	193485	1629 bp
191895	193485	1629 bp
191896	193485	1629 bp
191897	193485	1629 bp
191898	193485	1629 bp
191899	193485	1629 bp
191900	193485	1629 bp

Accession number	Contig of	bp in length
193980	198608	1876
194251	contig of	1876
196733	198608	1876

contig of	bp	in length
199109	1843	bp in length
200951	1621	bp in length
203072		
203453		

Contig	Length (bp)
203573	1478
205050	1478

Accession number	Contig length (bp)
205551	208808 contig of 1239 bp in length
207309	208969 contig of 1661 bp in length

Accession	Contig	Length (bp)
209470	210739	1270
209471	210739	1270
209472	210739	1270
209473	210739	1270
209474	210739	1270
209475	210739	1270
209476	210739	1270
209477	210739	1270
209478	210739	1270
209479	210739	1270
209480	210739	1270
209481	210739	1270
209482	210739	1270
209483	210739	1270
209484	210739	1270
209485	210739	1270
209486	210739	1270
209487	210739	1270
209488	210739	1270
209489	210739	1270
209490	210739	1270
209491	210739	1270
209492	210739	1270
209493	210739	1270
209494	210739	1270
209495	210739	1270
209496	210739	1270
209497	210739	1270
209498	210739	1270
209499	210739	1270
209500	210739	1270

213689		contig of	1048 bp in length.
214736			
215200			
215240			

* NOTE: This is a 'working draft' sequence. It currently contains 120 entries. The true order of the pieces is not yet known.

* is not known and their order in this sequence record is

* arbitrary. Gaps between the concepts are not
* arbitrary. Gaps between the concepts are not
* arbitrary. Gaps between the concepts are not

* This record will be updated with the finished sequence

* as soon as it is available.

* be preserved.

* 1 20626: contig of 20620 bp in tencan
* 20537 21120: gap of 503 bp

* 21130 32217: contig of 11088 bp in length

* 32216 32720: gap of 8506 bp in length

* 32721 41226: contig of 8506 bp in length

* 41227 41733: gap of 507 bp

* 50933 51432: gap of 500 bp.

* 51433 61043: colony of 5011 sp in 2009
* 61044 61545: gap of 502 bp

★ 61546 68789: contig of 7244 bp in length

* 69293 76356: contig of 7064 bp in length

* 76357 76858: gap of 302 bp
+ 76050 92702: contig of 6934 bp in length

* 83793 84296: gap of 504 bp
* 84297 90507: contig of 6211 bp in length
* 90508 91012: gap of 505 bp
* 91013 96837: contig of 5825 bp in length
* 96838 97337: gap of 500 bp
* 97338 102688: contig of 5351 bp in length
* 102689 103191: gap of 503 bp
* 103192 109162: contig of 5971 bp in length
* 109163 109664: gap of 502 bp
* 109665 114539: contig of 4875 bp in length
* 114540 115064: gap of 525 bp
* 115065 119306: contig of 4242 bp in length
* 119307 119814: gap of 508 bp
* 119815 125106: contig of 5292 bp in length
* 125107 125621: gap of 515 bp
* 125622 131031: contig of 5410 bp in length
* 131032 131531: gap of 500 bp
* 131532 137137: contig of 5606 bp in length
* 137138 137638: gap of 501 bp
* 137639 142257: contig of 4619 bp in length
* 142258 142758: gap of 501 bp
* 142759 147766: contig of 5008 bp in length
* 147767 148266: gap of 500 bp
* 148267 151608: contig of 3342 bp in length
* 151609 152108: gap of 500 bp
* 152109 155904: contig of 3796 bp in length
* 155905 156404: gap of 500 bp
* 156405 160842: contig of 4438 bp in length
* 160843 161342: gap of 500 bp
* 161343 164519: contig of 3177 bp in length
* 164520 165031: gap of 512 bp
* 165032 167901: contig of 2870 bp in length
* 167902 168402: gap of 501 bp
* 168403 172483: contig of 4081 bp in length
* 172484 172983: gap of 500 bp
* 172984 175602: contig of 2619 bp in length
* 175603 176104: gap of 502 bp
* 176105 179731: contig of 3627 bp in length
* 179732 180232: gap of 501 bp
* 180233 182324: contig of 2092 bp in length
* 182325 182831: gap of 507 bp
* 182832 186011: contig of 3180 bp in length
* 186012 186514: gap of 503 bp
* 186515 188995: contig of 2481 bp in length
* 188996 189495: gap of 500 bp
* 189496 191356: contig of 1861 bp in length
* 191357 191863: gap of 507 bp
* 191864 193485: contig of 1622 bp in length
* 193486 193985: gap of 500 bp
* 193986 196229: contig of 2244 bp in length
* 196230 196732: gap of 503 bp
* 196733 198608: contig of 1876 bp in length
* 198609 199108: gap of 500 bp
* 199109 200950: contig of 1842 bp in length
* 200951 201452: gap of 502 bp
* 201453 203072: contig of 1620 bp in length
* 203073 203575: gap of 503 bp
* 203576 205048: contig of 1473 bp in length
* 205049 205550: gap of 502 bp
* 205551 206808: contig of 1258 bp in length
* 206809 207308: gap of 500 bp
* 207309 208968: contig of 1660 bp in length
* 208969 209469: gap of 501 bp
* 209470 210739: contig of 1270 bp in length
* 210740 212339: gap of 500 bp
* 212340 213188: contig of 1949 bp in length
* 213189 213692: gap of 504 bp
* 213693 214736: contig of 1044 bp in length.

FEATURES
source
Location/Qualifiers
1. .214736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"

/clone="RP11-657A24"
/map="11q23"
BASE COUNT 50746 a 45972 c 45975 g 48371 t 23672 others
ORIGIN

Query Match 11.0%; Score 93.8; DB 32; Length 214736;
Best Local Similarity 63.3%; Pred. No. 1.8e-15;
Matches 193; Conservative 0; Mismatches 107; Indels 5; Gaps 3;

QY 545 ccacggcgattctccatctctcttgagatcaccatcgtgtgaagtcaggccacatgaggg 604
Db 118182 CCTGGGACATCCCTCTCTGGAACCCAGCTGCGATTCTGAGAAGCCCAAGCCACACAGAGT 118123
QY 605 tgcctgtgtagatgctccagctgaaatcccaagctaaagctcccaactgacagccaacat 664
Db 118122 GGCACATATAGGACACTCAAGTTGCCAGCACAGCTAAGCTCCACCACCAAGAGCCACCAT 118063
QY 665 -catttcagccatgtgtggagccatcctggatgtccagccttaacaaagccttcagagg 723
Db 118062 CCAGTCCAGCCATGTGAGTGAGCCATCTTGGATGTCCAGCTCAGCTGAGGCTTCAGACG 118003
QY 724 actcagccacagctattcttactacatccttctgtgaga-ctctaataaagaaccaact 782
Db 118002 ACTCCAGTCCAGCTGTGTGACTGCACACTCATGAGAAACAGACCCCAAGAGAATTTCCC 117943
QY 783 agctgagcccaatcaacctatggaactgatagataaaataaattgttttttcgaaa 842
Db 117942 AGCTGAGCCAGTCAACCCACAGAA---GTATGAGACATAATAAATGTTTTTAAAGATA 117886
QY 843 aaaaa 847
Db 117885 CTAAG 117881

RESULT 11
AC016336
LOCUS Homo sapiens clone RP11-31D17, WORKING DRAFT SEQUENCE, 14 unordered
DEFINITION pieces.
AC016336
AC016336.2 GI:6630581
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 129057)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-31D17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 129057)
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Dec 22, 1999 this sequence version replaced gi:6466997.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

```
* Contig_ID: 00815 Length: 1190bp
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```


Wed May 24 09:19:47 2000

SOURCE	human.	misc_feature	complement(16789..17120)
ORGANISM	Homo sapiens	repeat_region	/note="match: GSS A0002729 clone 49B17"
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	repeat_region	complement(17120..17294)
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.	repeat_region	/note="MERSA repeat: matches 189..6 of consensus"
TITLE	1 (bases 1 to 126352)	misc_feature	17359..17528
JOURNAL	McDougall,R.	repeat_region	/note="MIR repeat: matches 91..257 of consensus"
COMMENT	Direct Submission Submitted (08-JUL-1998) E-mail enquires: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 15, 1998 this sequence version replaced gi.2969933. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence is the entire insert of clone 23K20. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX 23K20 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR: pCYPAC2 .	repeat_region	/note="MIR repeat: matches 91..257 of consensus"
FEATURES	Location/Qualifiers	repeat_region	complement(17672..18039)
source	1..126352	repeat_region	/note="match: GSS B16679 clone A-343C02"
	/organism="Homo sapiens"	repeat_region	complement(18644..18768)
	/db_xref="taxon:9606"	repeat_region	/note="MIR repeat: matches 241..117 of consensus"
	/chromosome="X"	repeat_region	complement(18813..19103)
	/map="q25-26.2"	repeat_region	/note="AluSq repeat: matches 303..1 of consensus"
	/clone="RP1-23K20"	repeat_region	complement(19987..20100)
	/clone_lib="RPC1-1"	repeat_region	/note="MIR repeat: matches 237..113 of consensus"
	208..273	repeat_region	complement(20643..20824)
repeat_region	/note="MIR repeat: matches 124..189 of consensus"	repeat_region	/note="MIR repeat: matches 245..64 of consensus"
repeat_region	complement(710..1005)	repeat_region	complement(20886..21227)
repeat_region	/note="AluX repeat: matches 295..1 of consensus"	repeat_region	/note="MR42c repeat: matches 1538..1190 of consensus"
repeat_region	1174..1215	repeat_region	complement(21233..21587)
repeat_region	/note="21 copies 2 mer tg 93% conserved"	repeat_region	/note="THE1B repeat: matches 364..1 of consensus"
repeat_region	1420..1455	repeat_region	complement(21588..23146)
repeat_region	complement(2033..2159)	repeat_region	/note="THE1B-INTERNAL repeat: matches 1580..1 of consensus"
repeat_region	3965..4190	repeat_region	complement(23147..23503)
repeat_region	/note="MIR repeat: matches 21..262 of consensus"	repeat_region	/note="THE1B repeat: matches 364..1 of consensus"
repeat_region	complement(7046..7160)	repeat_region	complement(23610..23841)
repeat_region	/note="MIR2 repeat: matches 138..15 of consensus"	repeat_region	/note="L1ME1 repeat: matches 330..104 of consensus"
repeat_region	7689..7998	repeat_region	complement(24580..25191)
repeat_region	/note="AluJb repeat: matches 1..302 of consensus"	repeat_region	/note="L1 repeat: matches 4574..3960 of consensus"
repeat_region	complement(8180..8242)	repeat_region	25161..25537
repeat_region	/note="MIR2 repeat: matches 144..81 of consensus"	repeat_region	/note="THE1B repeat: matches 1..364 of consensus"
repeat_region	complement(8654..8949)	repeat_region	complement(25592..25734)
repeat_region	/note="AluSq repeat: matches 294..1 of consensus"	repeat_region	/note="L1 repeat: matches 3927..3781 of consensus"
repeat_region	complement(9170..9314)	repeat_region	complement(25846..26136)
repeat_region	/note="MIR2 repeat: matches 145..1 of consensus"	repeat_region	/note="AluJb repeat: matches 302..14 of consensus"
repeat_region	complement(13336..13644)	repeat_region	26948..27097
repeat_region	/note="AluX repeat: matches 303..1 of consensus"	repeat_region	/note="L1PA2 repeat: matches 744..893 of consensus"
repeat_region	complement(13810..13990)	repeat_region	complement(27070..27221)
repeat_region	/note="MIR repeat: matches 206..20 of consensus"	repeat_region	/note="match: STS L77624"
repeat_region	complement(14146..14450)	repeat_region	complement(27471..28299)
repeat_region	/note="AluX repeat: matches 302..1 of consensus"	repeat_region	/note="L1ME3 repeat: matches 910..62 of consensus"
repeat_region	14451..14515	repeat_region	complement(28310..28812)
repeat_region	/note="MIR repeat: matches 151..215 of consensus"	repeat_region	/note="L1ME1 repeat: matches 910..404 of consensus"
repeat_region	complement(14519..14631)	repeat_region	28899..29022
repeat_region	/note="MIR repeat: matches 167..48 of consensus"	repeat_region	/note="L1ME1 repeat: matches 399..521 of consensus"
repeat_region	16447..16509	repeat_region	29045..29094
repeat_region	/note="21 copies 3 mer cta 89% conserved"	repeat_region	/note="23 copies 2 mer ta 82% conserved"
repeat_region	complement(16558..16722)	repeat_region	complement(29106..29235)
	/note="MIR repeat: matches 215..57 of consensus"	repeat_region	/note="L1MD2 repeat: matches 419..290 of consensus"
		repeat_region	complement(29106..29274)
		repeat_region	/note="L1ME2 repeat: matches 419..247 of consensus"
		repeat_region	complement(29289..29591)
		repeat_region	/note="AluX repeat: matches 302..1 of consensus"
		repeat_region	complement(29622..29697)
		repeat_region	/note="L1 repeat: matches 5384..5306 of consensus"
		repeat_region	complement(30929..30983)
		repeat_region	/note="MIR repeat: matches 151..96 of consensus"
		repeat_region	31262..32179
		repeat_region	/note="L1 repeat: matches 2340..3253 of consensus"
		repeat_region	32191..32493
		repeat_region	/note="AluJb repeat: matches 1..302 of consensus"
		repeat_region	32494..33241
		repeat_region	/note="L1 repeat: matches 3263..4025 of consensus"
		repeat_region	33240..34062
		repeat_region	/note="L1MD1 repeat: matches 128..969 of consensus"
		repeat_region	35252..35649
		repeat_region	/note="MLT2FB repeat: matches 20..414 of consensus"
		repeat_region	35754..36196
		repeat_region	/note="MR42B repeat: matches 336..757 of consensus"
		repeat_region	36504..36806
		repeat_region	/note="AluSp repeat: matches 5..301 of consensus"
		repeat_region	36993..37290
		repeat_region	/note="AluY repeat: matches 2..301 of consensus"

```
unsure
repeat_region
/37270..37303
/37518..37603
/notes="MLT2FA repeat: matches 124..209 of consensus"
/complement(37694..37769)
/notes="MER45 repeat: matches 82..8 of consensus"
/37778..37866
/notes="L1 repeat: matches 3655..3744 of consensus"
/37871..38170
/notes="AluSg repeat: matches 1..294 of consensus"
/38172..38265
/notes="47 copies 2 mer ta 89% conserved"
/38374..38729
/notes="L1 repeat: matches 3852..4230 of consensus"
/39335..39423
/notes="L1 repeat: matches 4938..5031 of consensus"
/39489..39842
/notes="L1M4S repeat: matches 687..1045 of consensus"
/40077..40369
/notes="L1MB6 repeat: matches 9..323 of consensus"
/complement(40400..40921)
/notes="MLT2CA repeat: matches 493..1 of consensus"
/40927..41120
/notes="L1ME3 repeat: matches 358..555 of consensus"
/41111..41311
/notes="L1ME3 repeat: matches 708..901 of consensus"
/41427..41552
/notes="AluX repeat: matches 174..294 of consensus"
/42256..42592
/notes="MER2 repeat: matches 1..339 of consensus"
/complement(42896..43441)
/notes="L1 repeat: matches 1968..1427 of consensus"
/complement(43920..44981)
/notes="L1MA9 repeat: matches 1041..10 of consensus"
/complement(45231..45412)
/notes="AluJo repeat: matches 302..120 of consensus"
/complement(45423..45592)
/notes="MER42c repeat: matches 543..368 of consensus"
/complement(45670..45972)
/notes="MER42c repeat: matches 358..53 of consensus"
/46012..46377
/notes="MLT1A1 repeat: matches 4..365 of consensus"
/complement(47495..47635)
/notes="MIR repeat: matches 226..85 of consensus"
/complement(47882..48059)
/notes="MLT1F repeat: matches 536..359 of consensus"
/complement(48080..48341)
/notes="MLT1F repeat: matches 286..1 of consensus"
/complement(48363..48573)

Query Match      8.9%  Score 75.8;  DB 10;  Length 126352;
Best Local Similarity 62.5%  Pred. No. 1.8e-10;
Matches 168;  Conservative 0;  Mismatches 97;  Indels 4;  Gaps 3;

Qy 582 gtgaagtcagccagccatggaggtgctctgtgtagatgctccagctgaaatcccaagcta 641
Db 60481 GAGAAACCCACCTAAAAGAGCAAGACATGTAGTCTCCAGTCACACGCTCCTGATG 60422

Qy 642 agctcccaactgacagccaaatcattt-ccagccatgtgtggagccatcctggatgctc 700
Db 60421 AGCTCCAGCTGAATCTAACATCAACTGCCAGCTATGTAAAGTGAGCCATTTCAATGTT 60362

Qy 701 cagcttaacaagccttcagaggacttcagccacagctattattcttac--tacatccttg 758
Db 60361 TAGCCCGAGTTTGTCTCAGATGACTGTGATCTCCAGTTGCCCTCTGACTGTCACTGAATG 60302

Qy 759 tgagactctaataagaaccactagctgagcccaatcaacctatggaactgatagaaat 818
Db 60301 AAAGAGCAAAATTCGAACTTCCAGCTGAGCGCCGCCAAATCCACAGAACCATTTATAGAT 60242

Qy 819 aa-aatgaattgttttgcgaaaaaaa 846
Db 60241 AATAATGAATTGTGTTTTTAGCCACAAA 60213
```

RESULT 15
AC010404
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone C17B-H1_2149024, WORKING DRAFT
SEQUENCE, 42 unordered pieces.
AC010404
VERSION
AC010404.2 GI:6693230
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 220610)
DOE Joint Genome Institute.
AUTHORS
Sequencing of Human Chromosome 5
TITLE
Unpublished
JOURNAL
2 (bases 1 to 220610)
DOE Joint Genome Institute.
REFERENCE
Direct Submission
AUTHORS
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL
On Jan 14, 2000 this sequence version replaced gi:5882532.
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 139770 bases at least Q40
Consensus quality: 157240 bases at least Q30
Consensus quality: 165234 bases at least Q20
Estimated insert size: 220610; sum-of-contigs estimation
Estimated insert size: 140000; pulse field gel estimation
Quality coverage: 5.56x in Q20 bases; pulse field gel estimation
Quality coverage: 5.43x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1139: contig of 1139 bp in length
* gap of unknown length
* 1140 3706: contig of 2567 bp in length
* gap of unknown length
* 3707 6873: contig of 3167 bp in length
* gap of unknown length
* 6874 12010: contig of 5137 bp in length
* gap of unknown length
* 12011 19049: contig of 7039 bp in length
* gap of unknown length
* 19050 25097: contig of 6048 bp in length
* gap of unknown length
* 25098 26207: contig of 1110 bp in length
* gap of unknown length
* 26208 32540: contig of 6333 bp in length
* gap of unknown length
* 32541 35481: contig of 2941 bp in length
* gap of unknown length
* 35482 39346: contig of 3865 bp in length
* gap of unknown length
* 39347 40482: contig of 1136 bp in length
* gap of unknown length
* 40483 42160: contig of 1678 bp in length
* gap of unknown length
* 42161 46169: contig of 4009 bp in length
* gap of unknown length
* 46170 47438: contig of 1269 bp in length
* gap of unknown length
* 47439 48794: contig of 1356 bp in length

*	48795	49899:	gap of	unknown	length
*			contig of	1105 bp in	length
*	49900	51122:	gap of	unknown	length
*			contig of	1223 bp in	length
*	51123	52310:	gap of	unknown	length
*			contig of	1188 bp in	length
*	52311	53480:	gap of	unknown	length
*			contig of	1170 bp in	length
*	53481	54768:	gap of	unknown	length
*			contig of	1288 bp in	length
*	54769	55917:	gap of	unknown	length
*			contig of	1149 bp in	length
*	55918	57067:	gap of	unknown	length
*			contig of	1150 bp in	length
*	57068	58756:	gap of	unknown	length
*			contig of	1689 bp in	length
*	58757	59944:	gap of	unknown	length
*			contig of	1188 bp in	length
*	59945	61231:	gap of	unknown	length
*			contig of	1287 bp in	length
*	61232	62571:	gap of	unknown	length
*			contig of	1340 bp in	length
*	62572	63799:	gap of	unknown	length
*			contig of	1228 bp in	length
*	63800	65606:	gap of	unknown	length
*			contig of	1807 bp in	length
*	65607	66777:	gap of	unknown	length
*			contig of	1171 bp in	length
*	66778	68295:	gap of	unknown	length
*			contig of	1518 bp in	length
*	68296	69822:	gap of	unknown	length
*			contig of	1527 bp in	length
*	69823	72002:	gap of	unknown	length
*			contig of	2180 bp in	length
*	72003	73215:	gap of	unknown	length
*			contig of	1213 bp in	length
*	73216	75081:	gap of	unknown	length
*			contig of	1866 bp in	length
*	75082	77840:	gap of	unknown	length
*			contig of	2759 bp in	length
*	77841	80108:	gap of	unknown	length
*			contig of	2268 bp in	length
*	80109	85771:	gap of	unknown	length
*			contig of	5663 bp in	length
*	85772	94892:	gap of	unknown	length
*			contig of	9121 bp in	length
*	94893	107722:	gap of	unknown	length
*			contig of	12830 bp in	length
*	107723	120506:	gap of	unknown	length
*			contig of	12784 bp in	length
*	120507	141524:	gap of	unknown	length
*			contig of	21018 bp in	length
*	141525	220610:	gap of	unknown	length
*			contig of	79086 bp in	length

FEATURES
SOURCE

```
1. .220610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CITB-H1_2149024"
```

	BASE COUNT	ORIGIN	66468 a	42830 c	45423 g	61349 t	4540 others
/CROHE- CUBB IN-144024							

Query Match	8.9%	Score 75.6;	DB 44;	Length 220610;
Best Local Similarity	65.3%	Pred. No. 2.1e-10;		
Matches 111: Conservative	0;	Mismatches 59;	Indels	0; Gaps 0;

Qy 581 tgtgaagtcaggccacatggagggtgtcctgtgtagatgctccagctgaaatcccaagct 640
Dh 172450 tgcacaaattccaaacacatataaagcagccactgttagacacggcaggcagtagtgcacact 172518

Qy 641 aagctcccaactgacagagcccaacatcattccagccatgtgtggagccatcctggatgtc 700

[illegible]

Search completed: May 24, 2000, 00:59:16
Job time: 14307 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:02:30 ; Search time 282.38 seconds
(without alignments)
755.768 Million cell updates/sec

Title: US-09-215-435-123
Perfect score: 853
Sequence: 1 ggaggatggcgagcagctct.....tttgcgaaaaaaaaaaaaa 853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	189.8	22.3	239	X40653	Human secreted pro
2	170.4	20.0	296	V89317	EST clone C1437. N
3	91.4	10.7	184	X39568	Human secreted pro
c 4	34.4	4.0	110000	V121209_00	Methanococcus jann
5	33.6	3.9	13815	V18885	Mus musculus dystr
c 6	33.6	3.9	19307	T27558	Shuttle vector PAD
7	33	3.9	6407	T36035	rchd528 gene diffe
8	33	3.9	6407	V81825	Human rchd528 enco
9	33	3.9	6407	X26246	CDNA sequence of r
10	32.4	3.8	9060	V15702	Zucchini ACC synth
11	32.4	3.8	12923	N90338	Sequence of human
12	32.2	3.8	900	X34768	DNA encoding P30-8
c 13	31.8	3.7	858	Q55451	Pgs28 protein codl
c 14	31.8	3.7	858	Q87432	P. gallinaceum Pgs
c 15	31.6	3.7	1763	T17232	ACMNPV V-8 nt2469-
16	31.6	3.7	2133	T96754	S. tuberosum debra
c 17	31.2	3.7	110000	V21209_13	Continuation (14 o
c 18	31	3.6	2151	T33895	Lymphocyte chemoat
c 19	31	3.6	7515	X12865	Enterococcus faeca
c 20	31	3.6	9096	T88015	Human interleukin
c 21	31	3.6	19446	V52184	Streptococcus pneu
c 22	31	3.6	110000	V21209_03	Continuation (4 of
c 23	30.8	3.6	265	Q23880	pKTH1820 - promote
c 24	30.8	3.6	265	Q46205	Promoter sequence
c 25	30.8	3.6	265	T31875	Promoter clone PKT
c 26	30.8	3.6	588	V74487	Staphylococcus aur
c 27	30.8	3.6	2520	X08943	Fragment of human
c 28	30.8	3.6	9060	Q15132	Zucchini ACC synth
c 29	30.8	3.6	20633	X13213	Enterococcus faeca
c 30	30.8	3.6	110000	V21209_12	Continuation (13 o
c 31	30.6	3.6	532	V89371	EST clone CK234. N
c 32	30.4	3.6	3185	Q50430	Human MTP. Nucleic
c 33	30.4	3.6	110000	X20248_03	Continuation (4 of
c 34	30.2	3.5	1278	V59889	DNA encoding Staph

ALIGNMENTS

RESULT 1

X40653	ID	X40653 standard; cDNA; 239 BP.
AC	X40653;	
DE	18-JUN-1999	(first entry)
DE	Human secreted protein 5	EST SEQ ID No: 253.
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;	
KW	forensic; gene therapy; chromosome mapping; signal peptide; prostate;	
KW	upstream regulatory sequence; cytokine activity; cell proliferation;	
KW	differentiation; haematopoiesis regulation; tissue growth regulation;	
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;	
KW	thrombolytic; anti-inflammatory; tumour inhibition; ds.	
OS	Homo sapiens.	
PN	W09906550-A2.	
PD	11-FEB-1999.	
PF	31-JUL-1998; IB1232.	
PR	01-AUG-1997; US-905144.	
PA	(GEST) GENSET.	
PI	Ducier A, Dumas Milne Edwards J, Lacroix B;	
DR	WPI; 99-153780/13.	
DR	P-PSDB; Y11931.	
PT	New isolated prostate-derived nucleic acids - used to develop	
PT	products which may have cytokine, immune regulatory, haematopoiesis	
PT	regulating, anti-inflammatory or tumour inhibition activity	
PS	Claim 1; Page 398; 675pp; English.	
CC	X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human	
CC	secreted proteins expressed in prostate, and encode the proteins given in	
CC	Y11716 to Y11993 respectively. The proteins given represent the signal	
CC	peptide and an N-terminal fragment of a secreted protein. The nucleic	
CC	acid sequences can be used for producing secreted human gene products.	
CC	They can also be used to develop products for diagnosis and therapy. The	
CC	proteins obtained may have cytokine activity, cell proliferation and	
CC	differentiation activity, haematopoiesis regulating activity, tissue	
CC	growth regulating activity, reproductive hormone regulating activity,	
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,	
CC	receptor/ligand activity, anti-inflammatory activity, tumour inhibition	
CC	activity or other activities. The products can be used in forensic, gene	
CC	therapy and chromosome mapping procedures. The sequences can also be used	
CC	for obtaining corresponding promoter sequences. The nucleic acids	
CC	encoding the signal peptides can be used for directing extracellular	
CC	secretion of a polypeptide or the insertion of a polypeptide into a	
CC	membrane, or importing a polypeptide into a cell.	
SQ	Sequence, 239 BP; 63 A; 43 C; 67 G; 63 T;	

Query Match 22.3%; Score 189.8; DB 1; Length 239;
Best Local Similarity 99.0%; Pred. No. 2.8e-48;
Matches 189; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ggaggatggcgagcagctcgaatgccagaatggataaccgttttgctacagcattgta	60
Db	48	GGAGGATGGCGAGCAGTCTGNATGCCAGATGGTACCGTTTGTCTACAGCATTTGTA	107
Qy	61	attgcttggtgcttagcctcattttccaccattctacatggcagcctccattggcacagac	120
Db	108	ATDGTGTTGTGCTTAGCCTCATTTCCACCATCTACATGGCAGCCTCCATTGGCACAGAC	167

Staphylococcus aur
Streptococcus pneu
Human secreted pro
xpaGl/xpaG2 coding
Tobacco mosaic vir
Jaagsiekte retrovi
Carbamoyl-phosphat
Nucleotide sequenc
B61 cDNA sequence.
Human UCP3 gene tr
Human secretory pr

Wed May 24 09:19:48 2000

us-09-215-435-123.rng

QY 827 ttgtgttttgcgaataaa 850
 Db 8711 TCGTGAATTCAGAGAAGA 8688

RESULT 7
 T36035
 ID T36035 standard; DNA; 6407 BP.
 AC T36035; 1996 (first entry)
 DE rchd528 gene differentially expressed in cardiovascular disease.
 KW Cardiovascular disease; differential expression; target gene;
 KW pathway gene; fingerprint gene; atherosclerosis; ischaemia;
 KW reperfusion; hypertension; restenosis; arterial inflammation;
 KW vector; antibody; diagnosis; gene therapy; drug screening;
 KW rchd528 gene; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 19. .446
 FT /*tag= a

W09624604-A1.
 PD 15-AUG-1996.
 PF 09-FEB-1996; U01883.
 PR 10-FEB-1995; US-386844.
 PR 07-JUN-1995; US-485573.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Falb DA;
 DR WPI: 96-384391/38.
 DR P-PSDB; W03740.

PT New genes differentially expressed in cardiovascular disease - and related vectors, host cells, proteins and antibodies, for diagnosis, monitoring, treatment and drug screening
 PS Claim 1; Fig 31A-E; 200pp; English.
 CC 4 Novel genes, rchd502 (T36033), rchd523 (T36034), rchd528 (T36035) and rchd534 (T36036), are differentially expressed in endothelial cells subjected to shear stress. HUVEC cells were exposed to the laminar shear stress, which is thought to be responsible for the prevalence of atherosclerotic lesions in areas of unusual circulatory flow. RNAs from treated and control cells were used to generate cDNA libraries. Differentially expressed bands were identified by electrophoresis, subcloned and sequenced. The rchd528 gene product (W03740) was also sequenced. rchd528 is very highly expressed in the heart. 4 Other genes (see also T36033-36) are differentially expressed in endothelial cells exposed to interleukin-1.

CC Detection of these 8 novel genes in excess of normal levels allows the diagnosis of cardiovascular diseases (CVD). The genes can be used to generate diagnostic probes, to produce recombinant gene products, to breed transgenic animal models of CVD, in gene replacement therapy, or (as antisense, ribozyme or triplex sequences) to treat CVD.
 CC Sequence 6407 BP; 1592 A; 1763 C; 1523 G; 1507 T;
 SQ

Query Match 3.9%; Score 33; DB 1; Length 6407;
 Best Local Similarity 54.5%; Pred. No. 4.6;
 Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 86 ccaccatcacatggcagctccattggcacagactcttggtatgaatcgaatccag 145
 Db 2506 CCACCTTCTGCTCACAACACCTTGCACAGTCATCTTACCCCTGCTGCCAGG 2565
 QY 146 ttcaagaaattccagtgttgaataaaagcatctgggatgaatcattagatgaag 205
 Db 2566 GCAAGGAGACTCCTGTGACTTTCATTTTCAGACATCAACAATGACATTCATGACAATG 2625
 QY 206 c 206
 Db 2626 c 2626

RESULT 8
 V81825
 ID V81825 standard; cDNA; 6407 BP.
 AC V81825;

DT 11-MAR-1999 (first entry)
 DE Human rchd528 encoding cDNA.
 KW Human; cardiovascular disease; atherosclerosis; ischaemia; restenosis; reperfusion; hypertension; arterial inflammation; diagnosis; rchd528; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 19. .446
 FT /*tag= a
 FT US5849578-A.
 PD 15-DEC-1998.
 PF 15-MAR-1996; 616844.
 PR 09-FEB-1996; US-599654.
 PR 10-FEB-1995; US-386844.
 PR 07-JUN-1995; US-458873.
 PR 15-MAR-1996; US-616844.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Falb DA;
 DR WPI: 99-069743/06.
 DR P-PSDB; W89299.
 PT DNA encoding rchd528 polypeptide - associated with cardiovascular disease
 PS Claim 2; Fig 30; 122pp; English.
 CC The present sequence encodes rchd528 protein. A method has been developed for producing the rchd528 gene product. The present invention describes methods and compositions for the treatment and diagnosis of cardiovascular diseases, including: atherosclerosis; ischaemia; restenosis; reperfusion; hypertension; and arterial inflammation.
 CC Sequence 6407 BP; 1594 A; 1762 C; 1520 G; 1509 T;
 SQ

Query Match 3.9%; Score 33; DB 1; Length 6407;
 Best Local Similarity 54.5%; Pred. No. 4.6;
 Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 86 ccaccatcacatggcagctccattggcacagactcttggtatgaatcgaatccag 145
 Db 2506 CCACCTTCTGCTCACAACACCTTGCACAGTCATCTTACCCCTGCTGCCAGG 2565
 QY 146 ttcaagaaattccagtgttgaataaaagcatctgggatgaatcattagatgaag 205
 Db 2566 GCAAGGAGACTCCTGTGACTTTCATTTTCAGACATCAACAATGACATTCATGACAATG 2625
 QY 206 c 206
 Db 2626 c 2626

RESULT 9
 X26246
 ID X26246 standard; cDNA; 6407 BP.
 AC X26246;
 DE 24-MAY-1999 (first entry)
 DE cDNA sequence of rchd528 gene.
 KW Fingerprinting gene; rchd502; transmembrane protein; cardiovascular; fingerprint/target gene; up-regulated; endothelial cell; shear-stress; atherosclerosis; ischemia; reperfusion; hypertension; restenosis; ds.
 OS Homo sapiens.
 PN US5882925-A.
 PD 16-MAR-1999.
 PF 09-FEB-1996; 599654.
 PR 09-FEB-1996; US-599654.
 PR 10-FEB-1995; US-386844.
 PR 07-JUN-1995; US-485573.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Falb DA;
 DR WPI: 99-214071/18.
 DR P-PSDB; W95160.

PT New polynucleotides consisting of residues 1-1929 of the rchd502 gene - are differentially expressed in cardiovascular disease states, and can therefore be used to treat and diagnose cardiovascular diseases
 PS Disclosure; Fig 30A; 121pp; English.
 CC The invention relates to a rchd502 target/fingerprint gene encoding a


```

DR WPI: 94-035059/04.
DR P-PSDB; R48083.
PT Target antigens, esp. P28 for blocking transmission of antibodies
PT for malarial parasites - provides vaccine which induces high,
PT long-lasting antibodies and can be mass produced at low cost via
PT recombinant prodn.
PS Disclosure; Page 28-30; 42pp; English.
CC The Pgs28 protein and related proteins (Pfs28, Pfs25) can be used in
CC vaccines to prevent the transmission of malaria. Prevention of
CC transmission comprises administration of a pharmaceutical
CC composition comprising either a P28 protein or a recombinant virus
CC encoding a P28 protein in an amount sufficient to induce
CC transmission blocking immunity. The compositions may be used to
CC block the transmission of Plasmodium vivax, P. ovale, P. malariae and
CC P. falciparum. They may also be used to block transmission of P.
CC gallinaceum in chickens.
CC Sequence 858 BP; 319 A; 101 C; 145 G; 293 T;

Query Match 3.7%; Score 31.8; DB 1; Length 858;
Best Local Similarity 54.8%; Pred. No. 4.1;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 739 attattctactacatcctgtgtgagactctataaagaacaaactagctgagcccaatcaa 798
DB 255 ATTGCACTCAATATGATGCTCATCTCTATTAATAAACCATTTTACAAAGGTTTCATCAT 196
QY 799 cctatggaactgatagaataaataaataatgtgtgttgcgaaaaaataaaaaa 853
DB 195 CTGAAGGAGCTGCATAGTATATATATATGCAATTTGAATAAAGAAAGAAATA 141

RESULT 14
Q87432/c
ID Q87432 standard; DNA; 858 BP.
AC Q87432; 1995 (first entry)
DE P. gallinaceum Pgs28 gene.
KW Malaria; P28 protein; Pgs28; vaccine; immunogen;
OS transmission-blocking antibody; ss.
PT Plasmodium gallinaceum.
FH Key Location/Qualifiers
FT CDS 123..791
FT /*tag= a
FT WO9508631-A.
PD 30-MAR-1995.
PF 22-SEP-1994; U10720.
PR 22-SEP-1993; US-126593.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.
PI Duffy PE, Kaslow DC;
DR WPI; 95-139600/18.
DR P-PSDB; R71377.
PT Preventing transmission of malaria in susceptible organisms - by
PT administering a pharmaceutical composition comprising a Pfs28
PT polypeptide
PS Disclosure; Page 30-32; 38pp; English.
CC A genomic library of P. gallinaceum DNA was screened with probe
CC Nrl4AGT (given in Q87431) to identify colonies carrying the Pgs28
CC gene. Clone 9A1 was sequenced and found to have a 666 bp ORF
CC (Q87432) encoding Pgs28 (R71377).
CC Sequence 858 BP; 319 A; 101 C; 145 G; 293 T;

Query Match 3.7%; Score 31.8; DB 1; Length 858;
Best Local Similarity 54.8%; Pred. No. 4.1;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 739 attattctactacatcctgtgtgagactctataaagaacaaactagctgagcccaatcaa 798
DB 255 ATTGCACTCAATATGATGCTCATCTCTATTAATAAACCATTTTACAAAGGTTTCATCAT 196
QY 799 cctatggaactgatagaataaataaataatgtgtgttgcgaaaaaataaaaaa 853

```

```

DB 5850 AGTACAGAGCGAGGCTGATGATCTCTGAAATGCTTGATGACATTTGAAAAAATATAG 5909
QY 781 ctatgctgagcccaatcaactatggaactgatagaataaataaataatgtgttggcga 840
DB 5910 CCAGCTTACTTGAGCCAGAGATGAAGAAAGAAATTAAGGAATTTGATCGGGAATTCGAGA 5969
QY 841 aaaaaa 850
DB 5970 AGRAGAAAGA 5979

RESULT 12
X34768
ID X34768 standard; DNA; 900 BP.
AC X34768;
DT 05-JUL-1999 (first entry)
DE DNA encoding P30-8 protein.
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog; ss.
OS Ehrlichia canis.
PN WO9913720-A1.
PD 25-MAR-1999.
PF 18-SEP-1998; U19600.
PR 19-SEP-1997; US-059353.
PA (OHIS ) UNIV OHIO STATE.
PI Ohashi N, Rikihisa Y;
DR WPI; 99-254290/21.
DR P-PSDB; Y06968.
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
PS Disclosure; Fig 28A; 55pp; English.
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1A to -2) shown
CC in Y06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in Y06959-970. The proteins and genes are used
CC to detect E. chaffeensis in patients and E. canis in dogs.
CC Sequence 900 BP; 324 A; 142 C; 144 G; 290 T;

Query Match 3.8%; Score 32.2; DB 1; Length 900;
Best Local Similarity 57.4%; Pred. No. 3.2;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 149 aagaaataatccagtgattgataaagaacatctggatggaattcattagtgatgaagcag 208
DB 215 ATGTTAATCTTATTTCTATGAAACATCATGATATGTTGCTACAGGATTAAGCAAGCAACAA 274
QY 209 atgaaagactataatgatgacaccttttcgatacaatggc 249
DB 275 ATTTTAATCTTCTTATGTTGCAGAATTTCAAGACAATGCC 315

RESULT 13
Q55451/c
ID Q55451 standard; DNA; 858 BP.
AC Q55451;
DT 12-JUL-1994 (first entry)
DE Pgs28 protein coding sequence of Plasmodium gallinaceum.
KW Pgs28; Plasmodium vivax; Plasmodium ovale; Plasmodium malariae;
KW Plasmodium falciparum; Plasmodium gallinaceum; malaria; chicken;
KW transmission; prevention; vaccine; prophylaxis; ss.
OS Plasmodium gallinaceum.
FH Key Location/Qualifiers
FT CDS 123..788
FT /*tag= a
FT /product= Pgs28 protein.
FT WO9401552-A.
PD 20-JAN-1994.
PF 09-JUL-1993; U06464.
PR 10-JUL-1992; US-912294.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Duffy PE, Kaslow DC;

```

Db 195 CTGAAGGAGCTGCAATAGTCTAATATTGCAATTGCAATTAATAAAGAAAAATA 141

RESULT 15
T17232/c
ID T17232 standard; DNA; 1763 BP.
AC T17232;
DT 19-JUL-1996 (first entry)
DE ACMNPV V-8 nt2469-4231.
KW ACMNPV; baculovirus; insect pest; biological control; insecticide;
pesticide; late expression factor; lef-2; V1000 NPV; ss.
OS Autographa californica nuclear polyhedrosis virus strain V-8.
FH Key Location/Qualifiers
FT cds 194..826
FT /*tag= a
FT /label= lef-2
FT 3973..3814
FT /*tag= b
FT /label= ORF605
FT /note= "ORF605 is truncated in comparison to
the wild-type sequence"
FT cds 4135..4219
FT /*tag= c
FT /label= tolh
PN WO9603509-A2.
PD 08-FEB-1996.
PF 27-JUL-1995; U09527.
PR 27-JUL-1994; US-281916.
PA (AMCY) AMERICAN CYANAMID CO.
PI Ahmed F, Black BC, Dierks PM, Fleming NC, Miller LK;
DR WPI; 96-117054/12.
DR P-PSDB: R92770.
DT DNA molecule from Autographa californica nuclear polyhedrosis virus
for insect pests
PS Claim 2; Fig 4a-d; 65pp; English.
CC The MluI-EspI fragment (T17232) of the ACMNPV isolate V-8 genome
(nucleotides 2469-4219) includes the coding region for late
expression factor lef-2 (R92770). V-8 was obtd. from a viral clone
plaque purified on SF-21 cell monolayers inoculated with haemolymph
from Heliothis virescens larvae infected with Vail ACMNPV. It
shows an improved killing time for insect pests: the LT50 of
in Spodoptera frugiperda neonates was 12% shorter than the LT50 of
wild-type L-1. 72 Nucleotide differences and 4 insertions were
detected in the MluI-EspI fragment, predominantly in the
nt3002-4231 region, in comparison to the L-1 sequence (see also
T17231). It is postulated that V-8 is a recombinant between
ACMNPV and V1000 NPV. A further improvement in speed of killing
is obtd. by inactivation of the egt gene (see T17233). The
viruses can be used in the biological control of insect pests.
SQ Sequence 1763 BP; 582 A; 319 C; 301 G; 561 T;

Query Match 3.7%; Score 31.6; DB 1; Length 1763;
Best Local Similarity 60.5%; Pred. No. 6.6;
Matches 52; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 768 aataaagaaccaactagctgagcccaatcaacacctgtagaataaataaataa 827
Db 1139 AACAAAAAATTACTCTTTACGACAGATACATATATGATACGATAATAATATGCGAAT 1080
Qy 828 tgtgtgttttcgaaaaa 853
Db 1079 TTTTATGAAGAGAAAAATGAAAAAGA 1054

Search completed: May 24, 2000, 00:03:59
Job time: 10925 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:56:06 ; Search time 203.6 Seconds
(without alignments)
544,584 Million cell updates/sec

Title: US-09-215-435-123

Perfect score: 853

Sequence: 1 ggagatggggcagctctt.....tttgcgaaaaaaaaaaaaa 853

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.2	4.7	7218	1	US-08-232-463-14
c 2	33.6	3.9	19307	5	US-08-836-022A-10
3	33	3.9	6407	3	US-08-616-844-7
4	33	3.9	6407	3	US-08-599-654-7
5	33	3.9	6407	5	US-08-944-868A-7
6	33	3.9	6407	5	US-08-944-423A-7
c 7	32.4	3.8	7218	1	US-08-232-463-14
c 8	31.8	3.7	858	1	US-08-126-593A-1
c 9	31.8	3.7	858	2	US-08-454-039A-1
c 10	31.6	3.7	1763	1	US-08-281-916-3
c 11	31.6	3.7	1763	3	US-08-460-725-3
12	31.6	3.7	2625	5	US-08-804-439A-7
13	31.6	3.7	2625	5	US-08-720-229-7
c 14	31.2	3.7	845	4	US-08-743-637B-25
c 15	31.2	3.7	845	5	US-08-526-840B-25
c 16	31	3.6	2151	4	US-08-944-449-3
c 17	30.8	3.6	265	7	5242821-5
c 18	30.4	3.6	3185	1	US-08-117-362-2
c 19	30.4	3.6	3185	2	US-08-486-924-2
c 20	30.2	3.5	7400	1	US-08-261-663A-1
c 21	30.2	3.5	7400	6	PCT-US95-07754A-1
c 22	30.2	3.5	8920	3	US-08-446-855A-1
23	30	3.5	1498	1	US-08-321-162-1
24	30	3.5	1498	1	US-08-441-216-1
25	30	3.5	2000	3	US-08-948-277A-1
26	30	3.5	2000	4	US-09-169-203-1
27	30	3.5	3704	4	US-09-014-969-20
					Sequence 14, Appl
					Sequence 10, Appl
					Sequence 7, Appl
					Sequence 7, Appl
					Sequence 7, Appl
					Sequence 7, Appl
					Sequence 14, Appl
					Sequence 1, Appl
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 7, Appl
					Sequence 7, Appl
					Sequence 25, Appl
					Sequence 25, Appl
					Sequence 3, Appl
					Sequence 2, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 20, Appl

c 28 29.6 3.5 1196 3 US-08-864-799-2 Sequence 2, Appl
29 29.6 3.5 2223 2 US-08-257-073-4 Sequence 4, Appl
30 29.4 3.4 1586 1 US-07-841-646-18 Sequence 18, Appl
31 29.4 3.4 1586 1 US-08-147-023-18 Sequence 18, Appl
32 29.4 3.4 1586 2 US-08-447-570-18 Sequence 18, Appl
33 29.4 3.4 1586 3 US-08-449-700-18 Sequence 18, Appl
34 29.4 3.4 1586 4 US-08-449-699A-18 Sequence 3, Appl
35 29.4 3.4 1954 1 US-08-050-132A-3 Sequence 3, Appl
36 29.4 3.4 1954 3 US-07-989-847-3 Sequence 3, Appl
37 29.4 3.4 1954 5 US-08-750-222A-3 Sequence 3, Appl
38 29.4 3.4 1954 5 US-08-815-652B-3 Sequence 3, Appl
39 29.4 3.4 1954 6 PCT-US92-05374A-3 Sequence 3, Appl
40 29.4 3.4 1954 6 PCT-US95-07084-3 Patent No. 516058
41 29.4 3.4 1954 7 516058-5 Sequence 2, Appl
42 29.4 3.4 2148 5 US-08-809-999D-2 Sequence 2, Appl
43 29.4 3.4 2148 5 US-09-089-637-2 Sequence 2, Appl
c 44 29.4 3.4 2150 2 US-08-580-680-2 Sequence 2, Appl
c 45 29.4 3.4 2150 2 US-08-480-156A-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: ptzgt-F1s
: US-08-232-463-14

2566 GCAAGCCGACTCCTGACTTCATTTGAGACATCAACAATGACATCATTCATGACAATG 2625

Ov 206 c 206

Db 2626 C 2626

RESULT 6

```

US-08-944-423A-7
; Sequence 7, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
;
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITILE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; Zip: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,423A
; FILING DATE: 06-OCT-1997

```

Query Match	3.9%	Score 33;	DB 5;	Length 6407;
Best Local Similarity	54.5%	Pred. No. 1.4;		
Matches	66	Conservative	0;	Mismatches
			55;	Indels
			0;	Gaps

86 ccaccatctacatggcagccctccattggcacacagacttcttggtatgaatatcgaagtcag 145

2506 CCACTTTCTCACAACAACCTTGCCACAGTCATCTTCTACCCCTGTCTGCCCAGG 2565

146 ttcaagaaaaattccagttgatttgaataaaaqcatctcqqgatgaattcattagtgatgaag 205

2566 GCAAGGGAGACTCCTGTGACTTCATTTCAGACATCAACAATGACATCATTCATGACAATG 2625

206 c 206

Db 2626 C 2626

RESULT

```

US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-F1s
; US-08-232-463-14

```

Query Match 3.8%; Score 32.4; DB 1; Length 7218;
Best Local Similarity 2.0%; Pred. No. 2.3;
Matches 6. Conservative 166; Mismatches 122; Indels 0; Gaps 0;

122 +ctggtatgaatatcgaaagtccagttcagaagaaattccaqtgatttqaataaaagcatct 181

1413 TTTTCTACCCCCCCCCCCCCCCCCCC...
1383

182 **aggatgaattcattagtgatgaaacagatqaaaagacttataatgatgcaccttttcgat** 241

[illegible]

242 acaatggcacagtgqqqatttqtqqagacqqtatcaccatacccaaaacatgcattggt 301

Dh 1322 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1263

Qv 302 atagccaccagaaaqqacagagtcatttgatgtggtcacaaaatgtgtgagtttcacac 361

D6 1262 RR 1203


```

Query Match          3.7%; Score 31.6; DB 1; Length 1763;
Best Local Similarity 60.5%; Pred. No. 1.9; 34; Indels 0; Gaps 0;
Matches 52; Conservative 0; Mismatches 0;

Qy 768 aataagaaccaactagctgagcccaatcaacctatggaactgatagaataaaaaatgaat 827
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1139 AACAAAAAATTACTCTTTACGACAGATACATATNTGGATACCGATATATATTATGTGAAT 1080
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 828 tgttgttttcgaaaaaiaaaaaa 853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1079 TTTTATGAAGAGAAAATGAAAAGA 1054
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-460-725-3/c
; Sequence 3, Application US/08460725
; Patent No. 5858353
; GENERAL INFORMATION:
; APPLICANT: Miller, Lois K.
; APPLICANT: Black, Bruce C.
; APPLICANT: Dierks, Peter M.
; APPLICANT: Fleming, Nancy C.
; APPLICANT: Ahmed, Fakhruddin
; TITLE OF INVENTION: Insect Viruses, Sequences, Insecticidal
; TITLE OF INVENTION: Compositions and Methods
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder

```

```

RESULT 12
US-08-804-439A-7
; Sequence 7, Application US/08804439A
; Patent No. 6015565
; GENERAL INFORMATION:
; APPLICANT: Rose, Timothy M.
; APPLICANT: Bosch, Kurt L.
; APPLICANT: Strand, Martin
; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P. C.
; STREET: 4225 Executive Square, Ste 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804.439A
FILING DATE: February 21, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Haille, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09176/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2625 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-804-439A-7

Query Match 3.7%; Score 31.6; DB 5; Length 2625;
Best Local Similarity 54.2%; Pred. No. 2.4;
Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 276 caccatacccaaaacatgcttatagccaccagaaagacagagatcatttgatgt 335
Db 588 CGCCACCCCGAGAACCTGTGTGGTCTTTACACAAAGAACACAGTCAACTGCGAGGT 647

Qy 336 ggtcacaaaatgtgagtttcacactaactgagcagttcatggaattttggtat 393
Db 648 GACAGAGATGCTCGAGATCCATGAACCATTTTGAGTCTTTGTGACGCTCTGTTGGT 705

RESULT 13
US-08-720-229-7
Sequence 7, Application US/08720229
Patent No. 6022542
GENERAL INFORMATION:
APPLICANT: Rose, Timothy M.
APPLICANT: Bosch, Marnix L.
APPLICANT: Strand, Kurt
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
NUMBER OF SEQUENCES: 100
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/720.229
FILING DATE: 26-SEP-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Schliff, J. Michael
REGISTRATION NUMBER: 40,253
REFERENCE/DOCKET NUMBER: 29938-20002.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2625 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-720-229-7

Query Match 3.7%; Score 31.6; DB 5; Length 2625;
Best Local Similarity 54.2%; Pred. No. 2.4;
Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 276 caccatacccaaaacatgcttatagccaccagaaagacagagatcatttgatgt 335
Db 588 CGCCACCCCGAGAACCTGTGTGGTCTTTACACAAAGAACACAGTCAACTGCGAGGT 647

Qy 336 ggtcacaaaatgtgagtttcacactaactgagcagttcatggaattttggtat 393
Db 648 GACAGAGATGCTCGAGATCCATGAACCATTTTGAGTCTTTGTGACGCTCTGTTGGT 705

RESULT 14
US-08-743-637B-25/C
Sequence 25, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743.637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 845 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Haemophilus influenzae
US-08-743-637B-25

Query Match 3.7%; Score 31.2; DB 4; Length 845;
Best Local Similarity 55.6%; Pred. No. 1.7;

Wed May 24 09:19:48 2000

Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 743 tcttactacatccttgagactctaataagaacaaactagctgagcccaatcaacctta 802

Db 715 TATTGCTCAATTATTATGGATGGTAAATATCAGAAATATTGATGCATCAAGCATCTCC 656

QY 803 tggaaactgatagaataaataaatgaattgttttgcgaaaaaataaaa 850

Db 655 TGTAAATTATTAAATAAACTATTCTTTTTTTTACAAAAATAAAA 608

Search completed: May 23, 2000, 23:56:41
Job time: 10513 sec

RESULT 15

US-08-526-840B-25/c
; Sequence 25, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/526,840B
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Haemophilus influenzae
US-08-526-840B-25

Query Match 3.7%; Score 31.2; DB 5; Length 845;
Best Local Similarity 55.6%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 743 tcttactacatccttgagactctaataagaacaaactagctgagcccaatcaacctta 802

Db 715 TATTGCTCAATTATTATGGATGGTAAATATCAGAAATATTGATGCATCAAGCATCTCC 656

QY 803 tggaaactgatagaataaataaatgaattgttttgcgaaaaaataaaa 850

Db 655 TGTAAATTATTAAATAAACTATTCTTTTTTTTACAAAAATAAAA 608

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:02:32 ; Search time 3550.87 Seconds
(without alignments)
-1001.508 Million cell updates/sec

Title: US-09-215-435-123
Perfect score: 853
Sequence: 1 ggagtgatggcagcagtct.....tttgcgaaaaaaaaaaaaa 853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 5099991 seqs, -2084540658 residues 10199964
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA:*

1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq.*

45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US095_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq.*
86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq.*
87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq.*
88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq.*
90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq.*
91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq.*
92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq.*
93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq.*
94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq.*
95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq.*
96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq.*
97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq.*
98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	853	100.0	853	43	US-09-215-435-123
2	568.2	66.6	2039	56	US-09-471-275-850
3	566.8	66.4	1318	43	US-09-215-435-122
					Sequence 123, App
					Sequence 850, App
					Sequence 122, App

ALIGNMENTS

<hr/>	
RESULT 1	
US-09-215-435-123	
Sequence 123, Application US/09215435A	
GENERAL INFORMATION:	
APPLICANT: Dumas Milne Edwards, Jean-Baptiste	
APPLICANT: Duclert Aymeric,	
APPLICANT: Bougueleret Lydie	
TITLE OF INVENTION: Extended cDNAs	
FILE REFERENCE: GENSET.019A	
CURRENT APPLICATION NUMBER: US/09/215,435A	
CURRENT FILING DATE: 1998-12-17	
EARLIER APPLICATION NUMBER: 60/069,957	
EARLIER FILING DATE: 1997-12-17	
EARLIER APPLICATION NUMBER: 60/074,121	
EARLIER FILING DATE: 1998-2-9	
EARLIER APPLICATION NUMBER: 60/081,563	
EARLIER FILING DATE: 1998-4-13	
EARLIER APPLICATION NUMBER: 60/096,116	
EARLIER FILING DATE: 1998-8-10	
NUMBER OF SEQ ID NOS: 519	
SOFTWARE: Patent.pm	
SEQ ID NO 123	
LENGTH: 853	
TYPE: DNA	
ORGANISM: Homo sapiens	
FEATURE:	
<hr/>	

; CURRENT APPLICATION NUMBER: US/09/359,922
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1595
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-1595

Query Match 63.0%; Score 537.2; DB 51; Length 2076;
Best Local Similarity 98.1%; Pred. No. 4.5e-133;
Matches 565; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 1 ggaggatggcgagcagctgaatgccagatggataaccgtttgtctacagcatttgta 60
Db 101 ggaggatggcgagcagctgaatgccagatggataaccgtttgtctacagcatttgta 160

QY 61 atgctgtgtgtgtagcctcatttccaccatctacatggcagcctccattggcacagac 120
Db 161 atgctgtgtgtgtagcctcatttccaccatctacatggcagcctccattggcacagac 220

QY 121 ttctggtatgaatatcgaatccagttccagaaattccagtgatttgaataaaagcattc 180
Db 221 ttctggtatgaatatcgaatccagttccagaaattccagtgatttgaataaaagcattc 280

QY 181 tgggatgaattcattagtgatgaagcagatgaaagacttataatgatgcacctttt--c 238
Db 281 tgggatgaattcattagtgatgaagcagatgaaagacttataatgatgcaccttttcc 340

QY 239 gatacaatggcacagtgggattgtggagacggtgtatcaccataccaccaaaacatgcatt 298
Db 341 gatacaatggcacagtgggattgtggagacggtgtatcaccataccaccaaaacatgcatt 400

QY 299 ggtatagccaccagaaagacagagtcatttgatgtggtcacaaaatgtgtgagtttca 358
Db 401 ggtatagccaccagaaagacagagtcatttgatgtggtcacaaaatgtgtgagtttca 460

QY 359 cactaaactgagcagtcattgagaaattgttgcacggaacacacaaatagcggattg 418
Db 461 cactaaactgagcagtcattgagaaattgttgcacggaacacacaaatagcggattg 520

QY 419 atctccttaggacctatcttt--ggcgttggcagttcccttttacccttttggagtttaggt 477
Db 521 atctccttaggacctatctttggcgttggcagttcccttttacccttttggagtttaggt 580

QY 478 ttgatgtctttggggttggatcggactttgtcgttgcatcttccggaagcttatatccc 537
Db 581 ttgatgtctttggggttggatcggactttgtcgttgcatcttccggaagcttatatccc 640

QY 538 accattgccacgggcattctccatctccttgcagat 573
Db 641 accattgccacgggcattctccatctccttgcaggt 676

RESULT 6
US-09-359-922-1595
; Sequence 1595, Application US/09359922A
; GENERAL INFORMATION:
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Liu, Jin
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922A
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: US 09/034,341
; EARLIER FILING DATE: 1998-02-13

; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1595
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-1595

Query Match 63.0%; Score 537.2; DB 51; Length 2076;
Best Local Similarity 98.1%; Pred. No. 4.5e-133;
Matches 565; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 1 ggaggatggcgagcagctgaatgccagatggataaccgtttgtctacagcatttgta 60
Db 101 ggaggatggcgagcagctgaatgccagatggataaccgtttgtctacagcatttgta 160

QY 61 atgctgtgtgtgtagcctcatttccaccatctacatggcagcctccattggcacagac 120
Db 161 atgctgtgtgtgtagcctcatttccaccatctacatggcagcctccattggcacagac 220

QY 121 ttctggtatgaatatcgaatccagttccagaaattccagtgatttgaataaaagcattc 180
Db 221 ttctggtatgaatatcgaatccagttccagaaattccagtgatttgaataaaagcattc 280

QY 181 tgggatgaattcattagtgatgaagcagatgaaagacttataatgatgcacctttt--c 238
Db 281 tgggatgaattcattagtgatgaagcagatgaaagacttataatgatgcaccttttcc 340

QY 239 gatacaatggcacagtgggattgtggagacggtgtatcaccataccaccaaaacatgcatt 298
Db 341 gatacaatggcacagtgggattgtggagacggtgtatcaccataccaccaaaacatgcatt 400

QY 299 ggtatagccaccagaaagacagagtcatttgatgtggtcacaaaatgtgtgagtttca 358
Db 401 ggtatagccaccagaaagacagagtcatttgatgtggtcacaaaatgtgtgagtttca 460

QY 359 cactaaactgagcagtcattgagaaattgttgcacggaacacacaaatagcggattg 418
Db 461 cactaaactgagcagtcattgagaaattgttgcacggaacacacaaatagcggattg 520

QY 419 atctccttaggacctatcttt--ggcgttggcagttcccttttacccttttggagtttaggt 477
Db 521 atctccttaggacctatctttggcgttggcagttcccttttacccttttggagtttaggt 580

QY 478 ttgatgtctttggggttggatcggactttgtcgttgcatcttccggaagcttatatccc 537
Db 581 ttgatgtctttggggttggatcggactttgtcgttgcatcttccggaagcttatatccc 640

QY 538 accattgccacgggcattctccatctccttgcagat 573
Db 641 accattgccacgggcattctccatctccttgcaggt 676

RESULT 7
US-09-360-23786
; Sequence 23786, Application US/60172360
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE REFERENCE: GX-0007 P
; CURRENT APPLICATION NUMBER: US/60/172,360
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 29838
; SOFTWARE: PERL Program
; SEQ ID NO 23786
; LENGTH: 2127
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 403627.7
US-60-172-360-23786

Query Match      62.9%; Score 536.6; DB 92; Length 2127;
Best Local Similarity 98.1%; Pred. No. 6.6e-133; Indels 2; Gaps 2;
Matches 564; Conservative 0; Mismatches 9;

1 99agagatggcgagcagctgtaatgcagagtggaataaccgtttt-gctacagcatttgg 59
|||||
154 gtagcattctggagcagctgtaatgcagagtggaataaccgttttggctacagcatttgg 213
|||||
60 aattgctgtgtg-cttagcctcatttccaccatctacatggcagcctcattggcacag 118
|||||
214 aattgctgtgtgcttagcctcatttccaccatctacatggcagcctcattggcacag 273
|||||
119 acttctgtgtatgaatcgaagtcagcttcagaaataaccagtgatttgaataaaagca 178
|||||
274 acttctgtgtatgaatcgaagtcagcttcagaaataaccagtgatttgaataaaagca 333
|||||
179 tctgggatgaattcattagtgatgaagcagatgaaagacttataatgatgcacttttc 238
|||||
334 tctgggatgaattcattagtgatgaagcagatgaaagacttataatgatgcacttttc 393
|||||
239 gatacaatggcacagtggtgatttggagacggtgtatcacaccatacccaaaacatgcatt 298
|||||
394 gatacaatggcacagtggtgatttggagacggtgtatcacaccatacccaaaacatgcatt 453
|||||
299 ggtatagccaccagaaagacagacagtcatttgatggtgcacaaaatggtgagtttca 358
|||||
454 ggtatagccaccagaaagacagacagtcatttgatggtgcacaaaatggtgagtttca 513
|||||
359 cactaaactgagcagcttccatggagaaatttggatcccggaacacacacatagcggattg 418
|||||
514 cactaaactgagcagcttccatggagaaatttggatcccggaacacacacatagcggattg 573
|||||
419 atctccttagaacattcttggcgttgcagttccctttaccttttggatttagttt 478
|||||
574 atctccttagaacattcttggcgttgcagttccctttaccttttggatttagttt 633
|||||
479 tgatgtgttgggcttggatcggagacttggctgttcatttgcggaagcttatatccca 538
|||||
634 tgatgtgttgggcttggatcggagacttggctgttcatttgcggaagcttatatccca 693
|||||
539 ccattggcacgggacttccatctccttgcagat 573
|||||
694 ccattggcacgggacttccatctccttgcaggt 728
|||||

RESULT 8
US-60-070-771-1068
; Sequence 1068, Application US/60070771
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Au-Young, Janice
; APPLICANT: Guegler, Karl J.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: MICROARRAY FOR THE DETECTION OF
; TITLE OF INVENTION: EXPRESSION OF GENES CODING FOR SIGNAL SEQUENCE CONTAINING POLY
; NUMBER OF SEQUENCES: 1157
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/070,771
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0004 P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1068:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMN0T04
; CLONE: 2930673
US-60-070-771-1068

Query Match      61.6%; Score 525.6; DB 71; Length 1661;
Best Local Similarity 97.9%; Pred. No. 5.2e-130;
Matches 564; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

1 99agagatggcgagcagctgtaatgcagagtggaataaccgttttgcacagcatttga 60
|||||
111 GTAGCATCTCGAGCAGCTGTAATGCCAGATGGATAACCGTTTGTACAGCATTTGTA 170
|||||
61 attgctgtgtcttagcctcatttccaccatctacatggcagcctcattggcacagac 120
|||||
171 ATTGCTGTGTGCTTAGCTAGCTCAITTTCCACCATTCTACATGGCAGCCTCCATTTGGCAGACAC 230
|||||
121 tcttggtatgaatatcgaagtcagcttccaaagaaattccagtgatttgaataaaagcatc 180
|||||
231 TTCTGCTATGAATATCGAAGTCCAGTTCAGAAATTCAGTGATTGAATAAAAGCATC 290
|||||
181 tgggagtaattcattagtgatgaagcagatgaaagacttataatgatgcactttt-cg 239
|||||
291 TGGGATGAATTCATTAGTAGTGATGAGGAGATGAAAGACATTATAATGATGCATTTTTCGG 350
|||||
240 atacaatggcacagtggtgatttggagac-ggtgtatca-ccatcccccaaaacatgcatt 297
|||||
351 ATACAATGGCACAGTGGGATTGTGGAGACGGGTGTATCACCCCATACCCCAAAACATGCAT 410
|||||
298 tggatagccccaccagaaaggagcagctcatttggatggtgcacaaaatggtgagtttc 357
|||||
411 TGGTATAGCCCAACAGAAAGGACAGAGTCAATTTGATGTGGTGCACAAAATGTGTGAGTTTC 470
|||||
358 acactaaactgagcagcttccatggagaaatttggatcccggaacacacacatagcggatt 417
|||||
471 ACACTAACAGCAGCAGTTCATGGAGAAATTTGTTGATCCCGGAAACCAACAAATAGCGGATT 530
|||||
418 gatctccttaggacattcttggcgttgcaggttcccttttacccttttggatttaggt 477
|||||
531 GATCTCCTTAGGACCTATCTTTGGCGTGGCAGTTCTCTTTTACCTTTTGTGAGTTTAGGT 590
|||||
478 ttagtgcttgggcttggatcggagacttggctgttcatttgcagcagcttatatccc 537
|||||
591 TTGATGTGCTTTTGGGGCTTTTGATCGGACTTTTGTGCTTGGCATTTGCGGAAAGCTTATATCCC 650
|||||
538 accattggcacgggacttccatctccttgcagat 573
|||||
651 ACCATTGCCACGGGCATTTCTCCATCTCCTTGCAGGT 686
|||||

RESULT 9
US-60-164-285-415
```

```

; Sequence 415, Application US/60164285
; GENERAL INFORMATION:
; APPLICANT: Ma, Xiao-Jun
; TITLE OF INVENTION: Tumor Associated Molecules (TAMs): Targets for diagnosis, treatment
; FILE REFERENCE: 3214
; CURRENT APPLICATION NUMBER: US/60/164,285
; CURRENT FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 8259
; SEQ ID NO 415
; LENGTH: 2083
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-164-285-415

Query Match          61.6%; Score 525.6; DB 90; Length 2083;
Best Local Similarity 97.9%; Pred. No. 5.7e-130;
Matches 564; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

QY 1 ggaggatggcggagcagctgaatgccagaatggataaccgtttttgtacagcatttgta 60
   |||
Db 103 gtgcatcctggcagcagctgaatgccagaatggataaccgtttttgtacagcatttgta 162

QY 61 attgcttggtgtgttagcctcatttccacatctacatggcagcctccattggcacagac 120
   |||
Db 163 attgcttggtgtgttagcctcatttccacatctacatggcagcctccattggcacagac 222

QY 121 ttctggatgaatatcgaagtccagttccaagaaatcccagtgatttgaataaaagcattc 180
   |||
Db 223 ttctggatgaatatcgaagtccagttccaagaaatcccagtgatttgaataaaagcattc 282

QY 181 tgggatgaattcattagtgatgaagcagatgaaagacttataatgatgcacctttt-cg 239
   |||
Db 283 tgggatgaattcattagtgatgaggcagatgaaagacttataatgatgcaccttttccg 342

QY 240 atacaatggcacagtgggatgtgtggagac-ggtgtatcacattaccatacccaaaaacatgcatt 298
   |||
Db 343 atacaatggcacagtgggatgtgtggagacgggtgtatcacattacccaaaaacatgcatt 402

QY 299 ggtatagccacacagaaagacagagtcatttggatgggtcacaaaaatgtgtgagtttca 358
   |||
Db 403 ggtatagccacacagaaagacagagtcatttggatgggtcacaaaaatgtgtgagtttca 462

QY 359 cactaactgagcagttcatggagaaatttggatcccggaacacacaatagcgggattg 418
   |||
Db 463 cactaactgagcagttcatggagaaatttggatcccggaacacacaatagcgggattg 522

QY 419 atctccttagaacctatctttt-ggcgttgccagttccttttacccttttgtgagtttaggt 477
   |||
Db 523 atctccttagaacctatcttttggcgttgccagttccttttacccttttgtgagtttaggt 582

QY 478 ttgatgtccttggggcctttgatcgagacttgtgttcgatttgcctgaagccttatatccc 537
   |||
Db 583 ttgatgtccttggggcctttgatcgagacttgtgttcgatttgcctgaagccttatatccc 642

QY 538 accattgccacgggcattctccatctccttgcagat 573
   |||
Db 643 accattgccacgggcattctccatctccttgcaggt 678

RESULT 10
US-09-205-070-1401
; Sequence 1401, Application US/09205070
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS cDNA
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: 20411-748
; CURRENT APPLICATION NUMBER: US/09/205,070
; CURRENT FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1401

```



```

: TITLE OF INVENTION: Human Aortic Endothelium Library
:
: FILE REFERENCE: MLN98-14PA
:
: CURRENT APPLICATION NUMBER: US/09/307,140
:
: CURRENT FILING DATE: 1999-05-07
:
: EARLIER APPLICATION NUMBER: 60/084,565
:
: EARLIER FILING DATE: 1998-05-07
:
: NUMBER OF SEQ ID NOS: 1168
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ ID NO 72
:
: LENGTH: 586
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: (64)...(165)
:
: NAME/KEY: misc_feature
:
: LOCATION: (1)...(586)
:
: OTHER INFORMATION: n = A,T,C or G
:
: US-09-307-140-72

```

Query Match	53.3%	Score	455	DB	48	Length	590
Best Local Similarity	96.8%	Pred. No.	2.5e-111				
Matches	539	Conservative	0	Mismatches	7	Indels	11
Gaps							
Qy	1	ggaagatgggcgagcagctcgaatgccagaatgataaacgctttgtctacagcatttgtta	60				
Db	34	ggaagatgggcgagcagctcgaatgccagaatgataaacgctttgtctacagcatttgtta	93				
Qy	61	attgcttgtgtgttagcctcatttccaccatctacatggcagcctccattggcacagac	120				
Db	94	attgcttgtgtgttagcctcatttccaccatctacatggcagcctccattggcacagac	153				
Qy	121	tctcggta-tgaatatcgaatccagttccaagaataatccagtgatttgaat---aaaa	175				
Db	154	tctcggcatggatcgaagtcagttccaagaataatccagtgatttgaatpyraaaa	213				
Qy	176	gcattcgggatgaattcatttagtgatgaagcagatgaagaactataatgatgcacctt	235				
Db	214	gcattcgggatgaattcatttagtgatgagcagatgaagaactataatgatgcacctt	273				
Qy	236	ttcgatacaatggcacagtgggatttgtggaga-cgggtgatccacatacccaaaaacatg	294				
Db	274	ttcgatacaatggcacagtgggatttgtggacccggtgatccacatacccaaaaacatg	333				
Qy	295	cattggfatagcccacagaaaagcacagatcatttgaatggtgcacaaaaatggtgagt	354				
Db	334	cattggfatagcccacagaaaagcacagatcatttgaatggtgcacaaaaatggtgagt	393				
Qy	355	tccacataactgagcagttcatggagaaattt-gtgcaccggaaaccacaatagcgg	413				
Db	394	tccacataactgagcagttcatggagaaatttgggtgatcccggaaccacaatagcgg	453				
Qy	414	g-attgatctccttaggacctatatttggcgattgcacagttccttttacccttttgtgattt	472				
Db	454	gaattgatctccttaggacctatatttggcgattgcacagttccttttacccttttgtgattt	513				
Qy	473	taggtttt--gatgtgctttggggcttttgatcgagacttttgtgc-ttgcatttgcgcgaagt	529				
Db	514	taggttttggatgtgctttggggcttttgatcgagactttgtgcttgcatttgcgcgaagt	573				
Qy	530	tatatcccaacattgcc	546				
Db	574	tatatcccaacattgcc	590				

RESULT 14
US-09-289-768-26292
; Sequence 26292, Application US/09289768
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
OBTAINED

```

; FILE REFERENCE: 20411-765
; CURRENT APPLICATION NUMBER: US/09/289,768
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 39996
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26292
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(482)
; OTHER INFORMATION: n = A,T,C or G
US-09-289-768-26292

Query Match          52.5%; Score 447.8; DB 45; Length 482;
Best Local Similarity 99.6%; Pred. No. 1.9e-109;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps

QY 12 gacgagctgaatgccagaatgataaacgcttttgcgtacagcatttgcgttgatgtgt 71
    |||
Db 32 gacgagctgaatgccagaatgataaacgcttttgcgtacagcatttgcgttgatgtgt 91
    |||

QY 72 gcttagcctcatttcaccatctcacatgaggagcctccattggcacagactctgggtatga 131
    |||
Db 92 gcttagcctcatttcaccatctcacatgaggagcctccattggcacagactctgggtatga 151
    |||

QY 132 atatcgaagctccagttccaagaaattccagtgatttgaaataaagcatctgggataaatt 191
    |||
Db 152 atatcgaagctccagttccaagaaattccagtgatttgaaataaagcatctgggataaatt 211
    |||

QY 192 cattagtgatgaagcagatgaaagaaccttaaatgatgcaccttttcgatacaatggcac 251
    |||
Db 212 cattagtgatgaggcagatgaaagaaccttaaatgatgcaccttttcgatacaatggcac 271
    |||

QY 252 agtgggatttgagacggtgtatcaccatacccaaaaaacatgcattggtatagcccaacc 311
    |||
Db 272 agtgggatttgagacggtgtatcaccatacccaaaaaacatgcattggtatagcccaacc 331
    |||

QY 312 agaaaagcacagagctcattgatgtggtcacaaaatgtgtgagtttcacactaactgagca 371
    |||
Db 332 agaaaagcacagagctcattgatgtggtcacaaaatgtgtgagtttcacactaactgagca 391
    |||

QY 372 gtctcattggagaaatttggatcccggaaccacaatagcgggatttgatctctcttaggac 431
    |||
Db 392 gtctcattggagaaatttggatcccggaaccacaatagcgggatttgatctctcttaggac 451
    |||

QY 432 ctatctttggcgttgccagttctcttaacct 462
    |||
Db 452 ctatctttggcgttgccagttctcttaacct 482
    |||

```

```

RESULT 15
US-09-332-782-22673
; Sequence 22673, Application US/09332782A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/332,782A
; CURRENT FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22673
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(482)
; OTHER INFORMATION: n = A,T,C or G

```

US-09-332-782-22673

```
Query Match          52.5%; Score 447.8; DB 49; Length 482;
Best Local Similarity 99.6%; Pred. No. 1.1e-109;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 gagcagctgaatgccagaaatgataaccgttttgcacagcatttggtaattgcttctgtgt 71
Db 12 ||||||| 71
QY 32 gagcagctgaatgccagaaatgataaccgttttgcacagcatttggtaattgcttctgtgt 91
Db 32 ||||||| 91
QY 72 gcttagcctcatttccaccattcacatggcagcctccattggcagacacttctgggtatga 131
Db 72 ||||||| 131
QY 92 gcttagcctcatttccaccattcacatggcagcctccattggcagacacttctgggtatga 151
Db 92 ||||||| 151
QY 132 atatcgaagtcacagttcaagaataatccagtgatttgataataaagcactctgggatgaatt 191
Db 132 ||||||| 191
QY 152 atatcgaagtcacagttcaagaataatccagtgatttgataataaagcactctgggatgaatt 211
Db 152 ||||||| 211
QY 192 cattagtgatgaagcagatgaagaagacttataatgatgcaccttttcgatcacaaatggcac 251
Db 192 ||||||| 251
QY 212 cattagtgatgaagcagatgaagaagacttataatgatgcaccttttcgatcacaaatggcac 271
Db 212 ||||||| 271
QY 252 agtgggattgtggagacggtgtatcacccatacccaaaaacatgcattgggtatagccacc 311
Db 252 ||||||| 311
QY 272 agtgggattgtggagacggtgtatcacccatacccaaaaacatgcattgggtatagccacc 331
Db 272 ||||||| 331
QY 312 agaaagacagagtcatttgcattggtgcacaaaatgtgtgatttcacactaactgagca 371
Db 312 ||||||| 371
QY 332 agaaagacagagtcatttgcattggtgcacaaaatgtgtgatttcacactaactgagca 391
Db 332 ||||||| 391
QY 372 gttcatggagaaaattgtgatcccggaaccacaataagcgggattgatctccttaggac 431
Db 372 ||||||| 431
QY 392 gttcatggagaaaattgtgatcccggaaccacaataagcgggattgatctccttaggac 451
Db 392 ||||||| 451
QY 432 ctatcttggcgttggcagttccttttacct 462
Db 432 ||||||| 462
QY 452 ctatcttggcgttggcagttccttttacct 482
Db 452 ||||||| 482
```

Search completed: May 24, 2000, 01:02:48
Job time: 14340 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:59:16 ; Search time 6115.43 Seconds
(without alignments)
-131.393 Million cell updates/sec

Title: US-09-215-435-124
Perfect score: 826
Sequence: 1 aaccagagtgcccatgggt.....atcatcaaaaaaaaaaaa 826

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_p11:*
- 8: gb_p12:*
- 9: gb_prl:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sts:*
- 14: gb_sy:*
- 15: gb_un:*
- 16: gb_vi:*
- 17: em_fun:*
- 18: em_hum1:*
- 19: em_hum2:*
- 20: em_in:*
- 21: em_om:*
- 22: em_or:*
- 23: em_ov:*
- 24: em_pat:*
- 25: em_ph:*
- 26: em_pl:*
- 27: em_ro:*
- 28: em_sts:*
- 29: em_sy:*
- 30: em_un:*
- 31: em_vi:*
- 32: gb_htg1:*
- 33: gb_htg2:*
- 34: gb_in1:*
- 35: gb_in2:*
- 36: em_ba1:*
- 37: em_ba2:*
- 38: em_hum3:*
- 39: em_hum4:*
- 40: gb_pr4:*
- 41: gb_htg3:*
- 42: gb_htg4:*
- 43: gb_htg5:*
- 44: gb_htg6:*

- 45: gb_htg7:*
- 46: em_htg1:*
- 47: em_htg2:*
- 48: em_htg3:*
- 49: em_hum5:*
- 50: gb_pl3:*
- 51: gb_pr5:*
- 52: gb_htg8:*
- 53: gb_htg9:*
- 54: gb_htg10:*
- 55: gb_htg11:*
- 56: gb_htg12:*
- 57: gb_htg13:*
- 58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	347.8	42.1	352	13	G27363
2	47	5.7	1037	12	RNPBP
3	47	5.7	1047	5	E05646
4	47	5.7	1075	12	RNPEABP
C 5	43.2	5.2	82631	44	AC020316
6	43.2	5.2	92569	44	AC016445
7	41	5.0	564	24	E09005
8	41	5.0	570	10	HSRNAPBP
9	41	5.0	723	40	S76773
10	41	5.0	1001	24	E09006
11	41	5.0	1434	9	HMHRBPBP
12	41	5.0	1444	10	HSPEABP
13	41	5.0	1447	5	E05647
C 14	40.6	4.9	63551	42	AC012978
15	40.6	4.9	135417	32	HSDJ412I7
C 16	40.6	4.9	136933	35	AC005891
17	39.8	4.8	745	7	AB027456
18	39.8	4.8	194020	40	HUAC004125
19	39.6	4.8	528	8	AF152907
20	39.6	4.8	799	7	AB027506
21	39.4	4.8	1444	10	MFBBP
22	39.2	4.7	2441	10	HSTBX5
C 23	39.2	4.7	138383	40	AC005280
24	39.2	4.7	254129	44	AC009136
C 25	38.8	4.7	174138	52	AC011841
26	38.6	4.7	30334	42	AC013131
27	38.6	4.7	84055	35	AC004439
28	38.6	4.7	100000	9	AP000510
29	38.6	4.7	192650	9	AB023048
30	38.2	4.6	165393	44	AC016229
31	38	4.6	2104	5	A90842
32	38	4.6	152264	33	HSDJ493F7
33	38	4.6	168206	55	AC013792
C 34	37.8	4.6	179532	42	AC010856
35	37.6	4.6	44379	40	AC006163
C 36	37.6	4.6	146012	54	AC008651
C 37	37.4	4.5	145672	45	AC018771
38	37.4	4.5	160193	55	AC022576
C 39	37.4	4.5	185443	54	AC021310
40	37.2	4.5	239	10	HSPABD11
C 41	37.2	4.5	36542	40	AC006117
42	37.2	4.5	38852	40	AC006261
C 43	37.2	4.5	55369	45	AC010372
C 44	37.2	4.5	192262	43	AC011255
45	37	4.5	687	35	DMU05243

ALIGNMENTS

```
RESULT 1
G27363/c 352 bp DNA STS 28-JUN-1996
LOCUS human STS SHGC-31033, sequence tagged site.
DEFINITION G27363
ACCESSION G27363
VERSION 1 GI:1396086
KEYWORDS STS: STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 352)
AUTHORS Myers,R.M.
JOURNAL Unpublished (1996)
COMMENT Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu
Primer A: TTGGAAAGAGGGGTCTTG
Primer B: AAGCACAAAACCGGCG
STS size: 133
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds
Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
TaQ Polymerase: 0.05 units/ul
Total Vol: 10 ul
Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3
Prepared with primer pairs provided by Sandoz, derived from H83927
-- Washington University/Merck EST sequence.
FEATURES
source
1..352
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
4..136
Primer_bind 4..23
Primer_bind complement(119..136) 96 t 1 others
BASE COUNT 69 a 76 c 110 g 96 t
ORIGIN
Query Match 42.1%; Score 347.8; DB 13; Length 352;
Best Local Similarity 99.1%; Pred. No. 8.5e-86;
Matches 349; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 447 agtgggttcacgtaccagtttcttcttcacaggaaggtaacatctcttc 506
Db 352 AGTGGTTCATCGCTACCAAGTCTTGTCTATCTTCAGGAAGAAAGTCACTCTC 293
Qy 507 cttcccaaggaacaaactcgtcgttggaaaatggacagattttgacccgtttc 566
Db 292 CTTCCCAAGGAACAACTCGAGCTCTTGGAAATGGACAGATTCTCTGAACCGTTTC 233
```

```
Qy 567 caactggggcaacctgaagcaagcaccagttcatgaccagaactaccaggactcacca 626
Db 232 CACCTGGGCGAACTTGAAGCAAGCACCAGTTCATGACCCAGAACTTACCAGGACTCACCA 173
Qy 627 acctccaggtctccagagaaaggccagcagagcccaagacacaaaacccagcgagagata 586
Db 172 ACCTCCAGGCTCCAGAGAAAGGGCCAGGAGCCAGCCAGCAAAAACCCAGGCGGAGATA 113
Qy 687 gctgcctcgtagatagccggctttgcccattcggcattggtgcccacactgccaccaccga 746
Db 112 GCTGCCTGCTAGATAGNCCGCTTTGCCATCGGGCATGTGCCCACTGCCACACCCGA 53
Qy 747 cgatgtgggtatgaacccctctggtacagaaacccctcttctttccaata 798
Db 52 CGATGTGGGTATGGAACCCCTCTGGATACAGAACCCCTCTTTTCCAAATA 1
RESULT 2
RNPBP RNPBP 1037 bp mRNA ROD 17-FEB-1997
LOCUS R.norvegicus mRNA for phosphatidylethanolamine-binding protein.
DEFINITION X71873
ACCESSION X71873
VERSION X71873.1 GI:510338
KEYWORDS epididymal protein; pbp gene; phosphatidylethanolamine-binding protein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1037)
AUTHORS Perry,A.C., Hall,L., Bell,A.E. and Jones,R.
TITLE Sequence analysis of a mammalian phospholipid-binding protein from testis and epididymis and its distribution between spermatozoa and extracellular secretions
JOURNAL Biochem. J. 301 (Pt 1), 235-242 (1994)
MEDLINE 94311839
REFERENCE 2 (bases 1 to 1037)
AUTHORS Hall,L.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-1993) L. Hall, University of Bristol, Dept of Biochemistry, School of Medical Sciences, University Walk, Bristol BS8 1TD, UK
FEATURES
source
1..1037
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/dev_stage="adult male"
/tissue_type="epididymis and liver"
/clone_lib="cDNA"
/clone="18 clones"
1..28
5'UTR 1..28
gene 29..592
CDS 29..592
/gene="PBP"
/gene="PBP"
/codon_start=1
/product="phosphatidylethanolamine-binding protein"
/protein_id="CAA50708.1"
/db_xref="GI:510339"
/db_xref="SWISS-PROT:P31044"
/translation="NAADISQWAGPLSLQEYDEPPQHALRYDYGVTVDLGVLTPT
LYMNPSSISWDGLDPGLKLYTLVTDPAKRPKDPKFRWHHFLVYNNKMGDLSGTV
LSEYVSGSPKDTGLHRYVWLIVYEQQLNCDPELTLSKNKSDNRGKFKVESFRKKYHL
GARVAGTCFQAEWDDSVFKLHDLAGK"
590..1037
766
/notes="in 6 clones"
/replace="t"
BASE COUNT 244 a 299 c 290 g 204 t
ORIGIN
```



```
/clone_lib="lambda gt11"
/clone="pA01-1, pA01-10, pA62"
1. .1075
CDS
54. 617
/codon_start=1
/product="phosphatidyethanolamine binding protein"
/protein_id="CAA53032.1"
/db_xref="GI:406292"
/db_xref="SWISS-PROT:P31044"
/translation="MAADISQWAGPLSLQEVDEPPQHALRVYGGTVDELCKVLTP
QVNRPSISWGLDPKGKILVLIDPDAPSRKDFRWHFLVYVNMKGNLISGTV
LSEYVGGSPKDTGLHRYVWLYVEQEPQCDPELILSNKSGDNRGKFKVESPRKKYHL
GAPVAGTCFQAEMWDSVPRKLDQAGK"
polyA_signal
1015. .1020
polyA_signal
1036. .1075
BASE COUNT 261 a 305 c 298 g 211 t
ORIGIN
Query Match 5.7%; Score 47; DB 12; Length 1075;
Best Local Similarity 45.7%; Pred. No. 0.018;
Matches 164; Conservative 0; Mismatches 195; Indels 0; Gaps 0;
QY 281 ctatctctggtggtggtatccagatccctctagacagcagaacccagacagatt 340
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 242 CTACACCTGCTCTCACAGACCCCGATGCTCCAGCAGGAAGACCCCAATTCAGGGA 301
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 341 ctgagacattgctggttaacacatatcaagggcgccgacctgaagaagggaagattca 400
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 302 GTGGCACCACTCTCTGGTGTGAACATGAAGGGAACGACATTAGCAGTGGCACTGTCT 361
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 401 gggccaggattatcagctaccagggtccctccacacggccacacagtggtctccatcg 460
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 362 CTCGATACGTGGGCTCCGGACCTCCAAAGACACAGGCTCGCAGCGTACGCTGGCT 421
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 461 ctaccagttttgtctatcttcaggaaagggtatctctctcttcccaaggaaaa 520
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 422 GGTGTATGACGAGCAGCGCTCTGAACGTGACGAGCCCATCTCAGCAACAAGTCTGG 481
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 521 caaaactcaggctcttgaaaaagacagatttcgaacgtttccacctggcggaacc 580
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 482 AGACAACCGGGGAAGTCAAGTGGGAGTCCCTCCGCAAGAATACACCTGGGAGCCCC 541
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 581 tgaagcagcaccagtcattgaccagactaccaggactcaccacccctccaggctc 639
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 542 GGTGGCGGACGTGCTTCAGGCAGAGTGGGATGACTCTGTGCCCAAGCTGCATGATC 600
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
RESULT 5
AC020316/c
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
AC020316
AC020316.1 GI:6664581
VERSION
HTG; HTGS_PHASE2.
KEYWORDS
fruit fly.
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 82631)
Direct Submission
AUTHORS
Adams,M. and Venter,J.C.
TITLE
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
JOURNAL
This sequence was identified as CDM:10212745 by the submitter.
For more information on this record e-mail to fly@celera.com.
COMMENT
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
FEATURES
```

```
source
1. .82631
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 23613 a 17309 c 17873 g 23836 t
ORIGIN
Query Match 5.2%; Score 43.2; DB 44; Length 82631;
Best Local Similarity 50.5%; Pred. No. 0.24;
Matches 105; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 222 aagatcacctctggtgagcgcagatagtcagtctcccgggggccgtggacgcycaacc 281
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 72553 ACGCCACCCAGGTTAAGGATCATGCCAGTCGTGGAGTGGGACGCCGCGGGAGTTC 72494
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 282 tatatctcgtggtggtggtccagatgccccctagcagacagaaaccagacagattc 341
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 72493 TACACGCTATTATGACGATCCCGATGCCCGACGCGGACGCCCAAGTTCGCGCAG 72434
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 342 tggagacattgctggttaacacatatcaagggcgccgacctgaagaagggaagattcag 401
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 72433 TTCAACGACTGGATTCTGGCCACATCGCCGCGCAACGACTTGGCAAGTGGTAGCCCATC 72374
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 402 ggcacaggattatcagctaccaggctc 429
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 72373 GCGAGGTACATTGGCTCCGCGCCGCCGCC 72346
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
RESULT 6
AC016445
LOCUS
DEFINITION
Drosophila melanogaster chromosome 3 clone BACR19H03 (D1184)
RPT-98 19 H.3 map 83D-83F strain y; cn bw sp, *** SEQUENCING IN
PROGRESS ***, 69 unordered pieces.
AC016445
AC016445.3 GI:6633929
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
fruit fly.
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 92569)
Direct Submission
AUTHORS
Celnik,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Swirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
TITLE
Sequencing of Drosophila melanogaster
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 92569)
Celnik,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Swirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
JOURNAL
Submitted (30-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT
On Dec 23, 1999 this sequence version replaced gi:563465.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 268: contig of 268 bp in length
* 348: gap of unknown length
* 349: contig of 816 bp in length
* 1165: gap of unknown length
* 1245: contig of 520 bp in length
* 1844: gap of unknown length
* 1765: contig of 785 bp in length
* 1845: gap of unknown length
* 2630: contig of 735 bp in length
* 2710: gap of unknown length
* 3445: contig of 619 bp in length
* 3225: gap of unknown length
* 4144: contig of 552 bp in length
* 4224: gap of unknown length
* 4776: contig of 1133 bp in length
* 4856: gap of unknown length
* 5889: contig of 834 bp in length
* 6069: gap of unknown length
* 6903: contig of 704 bp in length
* 6983: gap of unknown length
* 7687: contig of 761 bp in length
* 7767: gap of unknown length
* 8527: contig of 516 bp in length
* 8607: gap of unknown length
* 8608: contig of 1189 bp in length
* 9123: gap of unknown length
* 9124: contig of 1108 bp in length
* 9204: gap of unknown length
* 10393: contig of 789 bp in length
* 10473: gap of unknown length
* 11581: contig of 1008 bp in length
* 11661: gap of unknown length
* 12450: contig of 981 bp in length
* 12539: gap of unknown length
* 13538: contig of 840 bp in length
* 13618: gap of unknown length
* 14599: contig of 840 bp in length
* 14679: gap of unknown length
* 15519: contig of 793 bp in length
* 15599: gap of unknown length
* 16392: contig of 817 bp in length
* 16472: gap of unknown length
* 17288: contig of 1060 bp in length
* 17369: gap of unknown length
* 17369: contig of 948 bp in length
* 18429: gap of unknown length
* 18509: contig of 828 bp in length
* 19457: gap of unknown length
* 19536: contig of 853 bp in length
* 20365: gap of unknown length
* 20445: contig of 760 bp in length
* 21298: gap of unknown length
* 21378: contig of 1133 bp in length
* 22138: gap of unknown length
* 22218: contig of 1591 bp in length
* 23351: gap of unknown length
* 23431: contig of 1939 bp in length
* 25022: gap of unknown length
* 25102: contig of 2226 bp in length
* 27041: gap of unknown length
* 27120: contig of 1363 bp in length
* 29347: gap of unknown length
* 29427: contig of 2090 bp in length
* 30790: gap of unknown length
* 30869: contig of 2474 bp in length
* 32960: gap of unknown length
* 32960: contig of 1276 bp in length
* 33040: gap of unknown length
* 35514: contig of 2505 bp in length
* 35594: gap of unknown length
* 36870: contig of 2505 bp in length
* 36950: gap of unknown length

* 39455: gap of unknown length
* 39535: contig of 1844 bp in length
* 41379: gap of unknown length
* 41458: contig of 2538 bp in length
* 43996: gap of unknown length
* 44076: contig of 2243 bp in length
* 44077: gap of unknown length
* 46320: contig of 3022 bp in length
* 46400: gap of unknown length
* 49422: contig of 2544 bp in length
* 49501: gap of unknown length
* 52045: contig of 3895 bp in length
* 52046: gap of unknown length
* 56021: contig of 2392 bp in length
* 56101: gap of unknown length
* 58493: contig of 2986 bp in length
* 58573: gap of unknown length
* 61559: contig of 4470 bp in length
* 61639: gap of unknown length
* 66109: contig of 2627 bp in length
* 66189: gap of unknown length
* 68816: contig of 3377 bp in length
* 68896: gap of unknown length
* 72273: contig of 5445 bp in length
* 72353: gap of unknown length
* 77798: contig of 497 bp in length
* 78375: gap of unknown length
* 78455: contig of 554 bp in length
* 79008: gap of unknown length
* 79089: contig of 563 bp in length
* 79652: gap of unknown length
* 79732: contig of 200 bp in length
* 79932: gap of unknown length
* 80012: contig of 435 bp in length
* 80447: gap of unknown length
* 80527: contig of 925 bp in length
* 81452: gap of unknown length
* 81532: contig of 663 bp in length
* 82195: gap of unknown length
* 82275: contig of 257 bp in length
* 82532: gap of unknown length
* 82612: contig of 609 bp in length
* 83221: gap of unknown length
* 83301: contig of 320 bp in length
* 83621: gap of unknown length
* 83701: contig of 888 bp in length
* 84589: gap of unknown length
* 84669: contig of 286 bp in length
* 84955: gap of unknown length
* 85035: contig of 454 bp in length
* 85489: gap of unknown length
* 85569: contig of 694 bp in length
* 86263: gap of unknown length
* 86343: contig of 411 bp in length
* 86754: gap of unknown length
* 86834: contig of 729 bp in length
* 87563: gap of unknown length
* 87643: contig of 1167 bp in length
* 88810: gap of unknown length
* 88890: contig of 554 bp in length
* 89444: gap of unknown length
* 89524: contig of 775 bp in length
* 90299: gap of unknown length
* 90379: contig of 606 bp in length
* 90985: gap of unknown length
* 91065: contig of 538 bp in length
* 91603: gap of unknown length
* 91683: contig of 523 bp in length
* 92206: gap of unknown length
* 92286: contig of 284 bp in length.

FEATURES
source
Location/Qualifiers
1. .92569
/organism="Drosophila melanogaster"

1


```

LOCUS       HUMRBPBP      1434 bp      mRNA      PRI      04-FEB-1999
DEFINITION   Human mRNA for human homologue of rat phosphatidyethanolamine
              binding protein, complete cds.
ACCESSION    D16111
VERSION      D16111.1  GI:435637
KEYWORDS     phosphatidylethanolamine binding protein homologue.
SOURCE       Homo sapiens cell-line HepG2 cDNA to mRNA, clone ha06-2.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1434)
AUTHORS      Hori,N.
TITLE        Direct Submission
JOURNAL      Submitted (26-APR-1993) to the DDBJ/EMBL/GenBank databases. Naohiro
              Hori, Osaka University, IMCB; Yamadaoka 1-3, Suita, Osaka 565,
              Japan (E-mail:hori@inherit.imcb.osaka-u.ac.jp,
              Tel:06-877-5111(ex.3314), Fax:06-877-1922)
REFERENCE    2 (bases 1 to 1434)
AUTHORS      Hori,N., Chae,K.S., Murakawa,K., Matoba,R., Fukushima,A., Okubo,K.
              and Matsubara,K.
TITLE        A human cDNA sequence homologue of bovine
              phosphatidylethanolamine-binding protein
JOURNAL      Gene 140 (2), 293-294 (1994)
MEDLINE      94193016
COMMENT      Submitted (26-APR-1993) to DDBJ by:
              Naohiro Hori
              Department of IMCB
              Osaka University
              1-3 Yamadaoka, Suita
              Osaka 565
              Japan
              Phone: 06-877-5111
              Email: hori@inherit.imcb.osaka-u.ac.jp
              Fax: 06-877-1922.

FEATURES             source
     source           1..1434
                        Location/Qualifiers
     CDS               92..655
                        /codon_start=1
                        /product="human homologue of rat phosphatidylethanolamine
                        binding protein"
                        /protein_id="BAA03684.1"
                        /db_xref="GI:435638"
                        /translation="MPVDLSKSGPLSLQVDPQPHLVTVAGAAVDELGKVLPTP
                        OVKNRPTISMDGLSDSKLYLTLPDAPSRKDPKYREWHFELVYVMKGNDISSGTV
                        LSDYVSGPPKGTGLHRYVWLVYEQDRPLKCDPEILLSRSGDHRGFKVASFRKKVEL"
BASE COUNT          350 a 343 c 398 g 343 t
ORIGIN

Query Match       5.0%; Score 41; DB 9; Length 1434;
Best Local Similarity 60.2%; Pred. No. 0.83;
Matches 68; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 281 ctatctcgtggtggtggtccagatgcccttagcagacagacagacagagatt 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 CTACACCTTGCTCTGACAGACCCGGATGCTCCAGCAGGAGGATCCCAATACAGAGA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 341 ctgagacattgctggttaacagatatcaaggcgccgacctgaagaaggga 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 ATGGCATCATTTCTGGTGGTCAACATGAAGGGCAATGACATCAGCTGGCA 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
HSPRBP
LOCUS       HSPRBP      1444 bp      mRNA      PRI      25-AUG-1995
DEFINITION   H.sapiens phosphatidylethanolamine binding protein mRNA.
ACCESSION    X75252
VERSION      X75252.1  GI:406289
KEYWORDS     phosphatidylethanolamine-binding protein.

```

SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS
Tohdoh,N., Tojo,S., Arai,H. and Ojika,K.

TITLE
Sequence homology of rat and human HCNP precursor proteins, bovine phosphatidylethanolamine-binding protein and rat 23-kDa protein associated with the opioid-binding protein

JOURNAL
MEDLINE
Brain Res. Mol. Brain Res. 30 (2), 381-384 (1995)

REFERENCE
MEDLINE
95364631

AUTHORS
Tohdoh,N.

TITLE
Direct Submission

JOURNAL
TITLE
Submitted (08-SEP-1993) N. Tohdoh, Sumitomo Pharmaceuticals Co., Ltd., 1-98, Kasugade-Naka 3-chome, Konohanaku, Osaka, 554, JAPAN

FEATURES
Location/Qualifiers
1..1444
/organism="Homo sapiens"
/db_xref="taxon:9606"
/germline
/tissue_type="placenta"
/clone_lib="lambda gtl1"
/clone="pp1-3, pp1-8"
111..674
/codon_start=1
/product="phosphatidylethanolamine binding protein"
/protein_id="CA53031.1"
/db_xref="GI:406290"
/translation="MPVDLSKWSGSLSQEVDEQPHLHVITYAGAAVDELGKVLTPPT QSKNRPSTISWDGLSDSGKLYTLVDPDAPSKRYREWHFLVVNMKGNDISGGT LSYDVCGSPKPGTGHLRVYLVVEODRPLKCDEPILSNRSGDHGRKFVKVAFPRKYEL RAPVAGTCQAEDWDDVPKLYLEQLSGK"

CDS
BASE COUNT 345 a 350 c 403 g 346 t
ORIGIN

Query Match 5.0%; Score 41; DB 10; Length 1444;
Best Local Similarity 60.2%; Pred. No. 0.83;
Matches 68; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 281 ctatatcctggtgatgtagtcagatgcacctagcagagcagaaccagacagagatt 340
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 CTACACCTTGCTCCTGCATGACAGCCGGATGTCCAGCAGGAAGGATCCCAATACAGAGA 358
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 341 ctggagacattgctgtatacacagatatcaaggcgccgacctgaagaaagga 393
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 ATGCATCAHTTCTGTTGGTCAACATGAAGGCCAATGACATCAGCAGTGCCA 411
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13

E05647
LOCUS
DEFINITION
E05647 1447 bp RNA PAT 29-SEP-1997
CDNA encoding human hippocampal cholinergic neurostimulating peptide.

ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens.

ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Tofuji,N., Tojo,S., Kojima,S., Ueki,Y., Nishihara,N., Fukushima,N., Irie,T., Ono,K., Agui,H. and Kojika,Y.,
NEUROTROPHIC FACTOR-RELATED POLYPEPTIDE AND GENE USEFUL THEREFOR
Patent: JP 1993268966-A 2 19-OCT-1993;

TITLE
JOURNAL
SUMITOMO PHARMACEUT CO LTD, KOJIKA YUKIO, YAMAMOTO MASAHICO

COMMENT
OS Homo sapiens (human)
PN JP 1993268966-A/2
PD 19-OCT-1993
PF 25-APR-1992 JP 1992131703


```

PR 27-APR-1991 JP 91P 124688
PI TOFUJI NAKI, TOJO SHINICHIRO, KOJIMA SHINICHI, UEKI YASUYUKI,
PI NISHIHARA NORIO, FUKUSHIMA NOBUYUKI, IRIE TSUNEMASA, PI ONO
KEIICHI,
PI AGUI HIDEO, KOJIKI YUKIO
PC C12N1/21, A61K37/02, A61K37/02, C07K7/06, C07K13/00, C12N1/21, PC
C12N5/10,
PC (C12N1/21, C12R1:19), (C12N5/10, C12R1:91);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=placenta;
CC *source: clone=pl-3;
FH Key Location/Qualifiers
FH 5'UTR <1..119
FH CDS 120..683
FT /product='hippocampal cholinergic FT
FT neurostimulating peptide'
FT 3'UTR 684..>1447.
FEATURES
    source
        1..1447
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 342 a 351 c 406 g 348 t
ORIGIN
Query Match 5.0%; Score 41; DB 5; Length 1447;
Best Local Similarity 60.2%; Pred. No. 0.83;
Matches 68; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 281 ctatctcgtgtggtggtatccagatgcccttagcagagcagacagacagagattt 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 308 CTACACCTTGTCTGCAGACACCGGATGCTCCACGAGGAGGATCCCAATACAGAGA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 341 ctggagacatggctggttaacagatatcaaggcgccgacctgaagaaggga 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 368 ATGGATCATTTCTCGGTGGTCAACATGAGGGCAATGACATCAGCAGTGGCA 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
AC012978/c
LOCUS AC012978 63551 bp DNA HTG 03-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC012978
VERSION AC012978.1 GI:6223345
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
    1 (bases 1 to 63551)
REFERENCE
    Adams.M. and Venter,J.C.
    Direct Submission
    Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
    Rockville, MD, USA
COMMENT
    This sequence was identified as CDM:10212016 by the submitter.
    For further information on this sequence you may e-mail to
    fly@celera.com.
    * NOTE: This is a 'working draft' sequence.
    * This sequence will be replaced
    * by the finished sequence as soon as it is available and
    * the accession number will be preserved.
FEATURES
    source
        1..63551
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
BASE COUNT 18144 a 13546 c 14099 g 17762 t
ORIGIN

```

```

Query Match 4.9%; Score 40.6; DB 42; Length 63551;
Best Local Similarity 53.5%; Pred. No. 1.3;
Matches 116; Conservative 0; Mismatches 89; Indels 12; Gaps 1;

QY 282 tatatcctgggtggtggtatccagatgcccttagcagagcagacccacagacagatttc 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22961 TACACCTCTGCATGACCGGATCGCGATCGGCCAGTCGCAAGGATCCCAAGTTAGGGAG 22902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 342 tggagacattggctggttaacagatatcaaggcgccgacctgaagaagggaagatttcag 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22901 TGGCACCATTGGCTGGTGGCAACATACCCGGTGGAGATGTGCCA-----AG 22854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 402 ggcagagattatcagcctaccaggtcctcccccacgcggcacacagtggttccatgc 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22853 GCGGAGGTTCTCTCCGCTACGTGGGATCCGGGCTCCACCAGACACCGGACTCCATCGT 22794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 462 taccagttcttctctctctcaggaagaaaggcca 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22793 TACGTTTCTCTGATCTACGACGCGGTGCAAGCTCA 22757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
HSDJ41217/c
LOCUS HSDJ41217 135417 bp DNA HTG 01-FEB-2000
DEFINITION Homo sapiens chromosome 6 clone RP3-41217 map q22.1-22.33, ***
SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION AL132795
VERSION AL132795.11 GI:6706641
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 135417)
REFERENCE
    Matthews.L.
    Direct Submission
    Submitted (14-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
    requests: clonerequest@sanger.ac.uk
    On Jan 16, 2000 this sequence version replaced gi:6688322.
    IMPORTANT: This sequence is unfinished and does not necessarily
    represent the correct sequence. Work on the sequence is in
    progress and the release of this data is based on the understanding
    that the sequence may change as work continues. The sequence may
    be contaminated with foreign sequence from E.coli, yeast, vector,
    phage etc. Order of segments is not known; 800 n's separate
    segments. Contig.ID: 02109 Length: 135417bp.
    * NOTE: This is a 'working draft' sequence.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
FEATURES
    source
        1..135417
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="q22.1-22.33"
            /clone="RP3-41217"
            /clone.lib="RP3-3"
BASE COUNT 43102 a 25039 c 25184 g 42092 t
ORIGIN
Query Match 4.9%; Score 40.6; DB 32; Length 135417;
Best Local Similarity 55.2%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 684 atagctgctgtagatagccggttggccatcccggtggtggtggtggtggtggtggtggt 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8677 ATTGCTTGACCCAGGAGCGGAGGTGTCAGTGAGCCAGAGATCGTGCCACTGCACCTCCAG 8618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: May 24, 2000, 01:02:06
Job time: 14477 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 00:03:59 ; Search time 282.38 seconds
(without alignments)
731.846 Million cell updates/sec

Title: US-09-215-435-124

Perfect score: 826

Sequence: 1 aaccagaggtgccatgggt.....atcatccaaaaa 826

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374.6	45.4	447	X40582	Human secreted pro
2	47	5.7	1047	Q30001	HCNP precursor gen
3	47	5.7	1047	Q58685	Rat hippocampal ch
4	41	5.0	1001	Q80734	Phosphatidylethano
5	41	5.0	1447	Q30002	HCNP precursor gen
6	41	5.0	1447	Q58686	Human hippocampal
7	38	4.6	2104	V40506	Human secreted pro
8	34.8	4.2	980	X37466	Human secreted pro
9	33.8	4.1	2182	V00731	Pea plastidial pho
10	33.6	4.1	4100	Q74062	The rat beta-actin
11	33.4	4.0	1032	Q23646	Acetylpolylamine am
12	33.4	4.0	1607	T85878	Malassezia fungus
13	33.2	4.0	852	N90094	ss cDNA number 19
14	33.2	4.0	852	N80721	Human SP5 #19. Rec
15	33.2	4.0	852	N70715	Sequence encoding
16	33.2	4.0	852	Q15263	SP-5 clone #19. Ne
17	33.2	4.0	858	N90099	ss cDNA number 18
18	33.2	4.0	858	N80641	Sequence of a comp
19	33.2	4.0	861	N80720	Human SP5 #18. Rec
20	33.2	4.0	861	N70714	Sequence encoding
21	33.2	4.0	861	Q15262	SP-5 clone #18. Ne
22	33.2	4.0	978	X51732	DNA encoding a hum
23	33.2	4.0	2483	V84468	Human secreted pro
24	32.8	4.0	745	V82783	Clone dn809_5 iso1
25	32.8	4.0	822	Q14832	OV-16 antigen. Onc
26	32.6	3.9	1733	V81394	Human tumour antig
27	32.4	3.9	1001	T33229	Wheat viviparous 1
28	32.4	3.9	3973	V65326	Human neuronatin g
29	32.2	3.9	913	T39051	cDNA encoding cell
30	32.2	3.9	940	T58281	Arabidopsis Strz po
31	32.2	3.9	1965	Q11815	HT7 membrane prote
32	32.2	3.9	2116	Q35987	Tomato hsp80 cDNA
33	32.2	3.9	4586	T96838	Intron 8 of human
34	32	3.9	2294	V79584	Rat organic anion

ALIGNMENTS

RESULT 1

X40582
ID X40582 standard; cDNA; 447 BP.

AC X40582;

DE 18-JUN-1999 (first entry)

KW Human secreted protein 5', EST SEQ ID No. 182.
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.

PN WO9906550-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; IB1232.

PR 01-AUG-1997; US-905144.

PA (GEST) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI; 99-153780/13.

DR P-PSDB; Y11860.

PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity

PS Claim 1; Page 285; 675pp; English.

CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins expressed in prostate, and encode the proteins given in
CC X11716 to Y11993 respectively. The proteins given represent the signal
CC peptide and an N-terminal fragment of a secreted protein. The nucleic
CC acid sequences can be used for producing secreted human gene products.

CC They can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell proliferation and
CC differentiation activity, haematopoiesis regulating activity, tissue
CC growth regulating activity, reproductive hormone regulating activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,

CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC activity or other activities. The products can be used in forensic, gene
CC therapy and chromosome mapping procedures. The sequences can also be used
CC for obtaining corresponding promoter sequences. The nucleic acids
CC encoding the signal peptides can be used for directing extracellular

CC secretion of a polypeptide or the insertion of a polypeptide into a
CC membrane, or importing a polypeptide into a cell.
SQ Sequence 447 BP; 98 A; 110 C; 136 G; 99 T;

Query Match

45.4%; Score 374.6; DB 1; Length 447;

Best Local Similarity 99.5%; Pred. No. 7.7e-102;

Matches 375; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 aaccagaggtgccatgggtggacaatgagcctgggtcacagcagcactgtactgggtc 60

|||||

Db 71 AACACAGAGGTGCCATGGGTGGACATGAGCTGGTGGTCACAGACGACCTGTATTGGGTC 130

Qy 61 tcacatggtggtcactggagacgagatgatgaacagcccgctgtgccatgagccctcc 120

|||||

Db 131 TCATGATGGTGGTCTACATGGACGACGAGGATGAGACAGCCCGTGTGCCATGAGGCCCTCT 190

	QY	281	ctatatcctgggtgatggttgatccagatgccctccctacagacagaaaccacacagagatt	340
	Db	214	CTACACCCTGGTCTCTCACAGCCCCGATGCTCCACAGCAAGAGGCCCCCAAATTTCAGGA	273
	QY	341	ctggagacattggctggtaaacagatatcaaggcgccgcacctgaagaagaaggaattcca	400
	Db	274	GTGGCACCACCTTCTGGTGTTCAACATFGAAGGCACGACANTTAGCAGTGGCAGCTCTCT	333
	QY	401	ggggccaggagtattacagctaccaggctcccctccccaccggcacacagttggcttccatcg	460
	Db	334	CTCCGAATACGTGGGCTCGGAGCTCCCAAAGACACAGGCTGTGCACCGCTACGCTCTGGCT	393
	QY	461	ctaccagttcttgtctatcttcagaaagaaaaggtcatctctctctcccaaggaaaa	520
	Db	394	GGTGTATGACACAGGACAGCCTCTGAACATGTGACGAGCCCATCTCTCAGACAAGAAGTCTGG	453
	QY	521	caaaactcggaggtctcttgaaaatggacagattcttgaaccgtttccacctggggqaacc	580
	Db	454	AGACAACCGCGCAAGTTCAAGGTGGAGTCTTTCGCAAGAAGTACCACTGGGAGCCCC	513
	QY	581	tgaagcaagaccacaggttatgaccagacaactaccaggactcacaacccctccaggctc	639
	Db	514	GGTGGCGGCACGCTGTTCCAGCAGAGTGGGATGACTCTGTGCCCAAGCTGCATGATC	572
	RESULT	4		
	Q80734			
	ID	Q80734	standard; cDNA to mRNA; 1001 BP.	
	AC	Q80734;		
	DT	04-JUL-1995	(first entry)	
	DE	phosphatidylethanolamine binding protein coding sequence.		
	KW	Human; phosphatidylethanolamine binding protein; hPEBP; cDNA library;		
	KW	glioblastoma; T986; aplasia; proliferation; glia; neuron;		
	KW	hematopoietic cell; immunological activity; neurological activity;		
	KW	inflammatory disease; tumour; antibody; ss.		
	OS	Homo sapiens.		
	FH	Key	Location/Qualifiers	
	FT	cds	53..616	
	FT		/tag= a	
	FT		/product= hPEBP	
	PN	EP-628631-A.		
	PD	14-DEC-1994.		
	PF	30-MAY-1994; 108321.		
	PR	08-JUN-1993; JP-137042.		
	PA	(ONOX) ONO PHARM CO LTD.		
	PI	Naito T, Nakade S;		
	DR	WPI; 95-015698/03.		
	DR	P-PSDB; R64268.		
	PT	New human phosphatidylethanolamine binding protein gene - used to		
	PT	develop prods. for diagnosis and treatment of diseases involving		
	PT	glia, neurons and other cells.		
	PS	Claim 4; Page 13-14; 17pp; English.		
	CC	This sequence encodes the human phosphatidylethanolamine binding		
	CC	protein, hPEBP. This cDNA was isolated from a cDNA library prepared		
	CC	using RNA extracted from the human glioblastoma cell line T986. The		
	CC	protein can be used for the prevention of or in the treatment of		
	CC	aplasia or abnormal proliferation of glia, neurons or hematopoietic		
	CC	cells, depression or enhancement of immunological or neurological		
	CC	activity, inflammatory disease, tumours or diseases induced by		
	CC	abnormal lipid metabolism. Anti-hPEBP antibodies may be used for		
	CC	diagnosis of genetic diseases.		
	SQ	Sequence 1001 BP; 250 A;		
		C 249 C; 279 G; 223 T;		

Db	301	ATGGCATCATTTCTGGTGGTCAACATGAAGGCAATGACATCATGACGAGTGGCA	353
Db	301	ATGGCATCATTTCTGGTGGTCAACATGAAGGCAATGACATCATGACGAGTGGCA	353
RESULT	5		
Q30002			
ID	Q30002	standard; cDNA to mRNA; 1447 BP.	
AC	Q30002;		
DE	18-MAR-1993	(first entry)	
DE	HCNP precursor gene #2.		
KW	Hippocampal cholinergic neurotrophic peptide; HCNP; hippocampal tissue;		
KW	expression vector; neurodegenerative disorder; dementia;		
KW	Alzheimer's disease; Parkinson's disease; ss.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	cds	120..680	
FT		/*tag= a	
FT	5'utr	1..119	
FT		/*tag= b	
FT	3'utr	681..1447	
FT		/*tag= c	
FT	misc_rna	123..155	
FT		/*tag= d	
FT		/label= HCNP	
FN	EP-511816-A.		
PD	04-NOV-1992.		
PF	27-APR-1992; 303800.		
PR	27-APR-1991: JP-124688.		
PA	(OJIK)/ OJIK K. OJIK A. CO LTD.		
PA	(SUMO) SUMITOMO PHARM CO LTD.		
PA	(YAMA) YAMAMOTO M.		
PI	Aqui H, Fukushima N, Irie T, Kojima S, Nishihara T;		
PI	Ojika K, Ono K, Tohdoh N, Tojo S, Ueki Y;		
DR	WPI: 92-367633/45.		
DR	P-PSDB; R27718.		
PT	Neurotrophic peptide derivs. - used for treating neurological		
PT	degenerative disorders, e.g. Alzheimer's disease or Parkinson's		
PT	disease		
PS	Disclosure: Page 42-44; 57pp; English.		
CC	The sequence given is the human hippocampal cholinergic neurotrophic		
CC	peptide (HCNP) precursor gene. The active part of the protein encoded		
CC	by this sequence is located in the first eleven amino acids at the N		
CC	terminus. This gene was isolated using the rat HCNP precursor gene		
CC	sequence as a guide (see also Q30001). From studies of the		
CC	structure of the rat HCNP precursor gene and protein the human HCNP		
CC	was determined. The rat and human HCNP's have markedly different		
CC	structures and composition. HCNP's are useful for the treatment of		
CC	neurodegenerative disorders and dementia eg. Alzheimer's and		
CC	Parkinson's disease.		
SQ	Sequence 1447 BP; 342 A; 351 C; 406 G; 348 T;		
Query Match	5.0%;	Score 41; DB 1; Length 1447;	
Best Local Similarity	60.2%;	Pred. No. 0.015;	
Matches	68; Conservative	0; Mismatches 45; Indels 0; Gaps 0;	
Qy	281	ctatctcgtgatggtggtccagatgcccttagcagagcagacacacagagatt	340
Db	308	CTACACCTTGGTCTGACAGACCCCGATGCTCCACGACGAGGAGGATCCCAATACAGAGA	367
Qy	341	ctggagacattggctggtaacagatatcaaggccgacacctgaagaaggga	393
Db	368	ATGGCATCATTTCTGGTGGTCAACATGAAGGCAATGACATCATGACGAGTGGCA	420
RESULT	6		
Q58686			
ID	Q58686	standard; cDNA to mRNA; 1447 BP.	
AC	Q58686;		
DT	13-OCT-1994	(first entry)	
DE	Human hippocampal cholinergic neurotrophic peptide precursor cDNA.		
KW	Human hippocampal cholinergic neurotrophic peptide; human HCNP;		
KW	nerve degeneration; acetylcholine synthesis; neurostimulation; ds.		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 23:56:41 ; Search time 203.6 seconds
(without alignments)
527.346 Million cell updates/sec

Title: US-09-215-435-124
Perfect score: 826
Sequence: 1 aaccagagtgccatgggt.....atcatcaaaaaaaaaaaaaa 826

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues
Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	45.4	5.5	1047	2 US-08-403-378B-3
2	41	5.0	1447	2 US-08-403-378B-14
3	35	4.2	7218	1 US-08-232-463-14
4	32.8	4.0	822	1 US-07-644-372-1
5	32.4	3.9	3973	3 US-08-602-093-6
6	32.2	3.9	913	5 US-08-651-136C-13
7	32.2	3.9	940	3 US-08-471-717-1
8	32.2	3.9	4586	2 US-08-832-883-53
9	32.2	3.9	4586	3 US-08-832-877-53
10	31.6	3.8	1558	2 US-08-467-607-2
11	31.6	3.8	1558	3 US-08-469-362-2
12	31.6	3.8	1558	3 US-08-850-392-2
13	31.6	3.8	1641	1 US-08-300-903A-8
14	31.6	3.8	1712	3 US-08-632-598-1
15	31.6	3.8	3581	3 US-08-738-349-1
16	31.2	3.8	1338	1 US-08-631-200-14
17	31.2	3.8	1338	2 US-08-829-553-14
18	31.2	3.8	1338	3 US-08-922-267A-14
19	31.2	3.8	1338	3 US-08-936-707A-14
20	31.2	3.8	1338	3 US-08-936-706A-14
21	31.2	3.8	1338	5 US-09-248-203-14
22	31.2	3.8	1642	3 US-08-665-037-1
23	31.2	3.8	1642	4 US-08-666-067-1
24	31.2	3.8	1642	4 US-08-732-870-1
25	31.2	3.8	1659	2 US-08-548-509-4
26	31.2	3.8	1687	1 US-08-143-219-26
27	31	3.8	1341	4 US-08-997-080-93

c	28	31	3.8	1341	4	US-08-997-362-93	Sequence 93, Appl
c	29	31	3.8	1341	5	US-08-873-970-93	Sequence 93, Appl
	30	31	3.8	2599	7	5266464-1	Patent No. 5266464
	31	30.2	3.7	1101	5	US-08-987-904A-1	Sequence 1, Appl
	32	30.2	3.7	1174	4	US-08-872-437-1	Sequence 1, Appl
	33	30.2	3.7	1174	5	US-08-651-136C-11	Sequence 11, Appl
c	34	30.2	3.7	32207	3	US-08-770-379-20	Sequence 12, Appl
c	35	30	3.6	70	4	US-08-776-944-12	Sequence 20, Appl
	36	30	3.6	1654	5	US-08-991-426-3	Sequence 3, Appl
	37	30	3.6	1654	5	US-09-143-470-3	Sequence 3, Appl
c	38	30	3.6	2040	3	US-08-533-669A-5	Sequence 5, Appl
	39	30	3.6	4104	1	US-07-998-003A-94	Sequence 94, Appl
	40	30	3.6	4104	1	US-08-453-274B-94	Sequence 94, Appl
	41	30	3.6	4104	2	US-08-453-695A-94	Sequence 94, Appl
	42	30	3.6	4104	2	US-08-268-161A-94	Sequence 94, Appl
	43	30	3.6	4104	3	US-08-453-702A-94	Sequence 94, Appl
	44	30	3.6	4104	6	PCT-US93-12588-94	Sequence 94, Appl
	45	30	3.6	4104	6	PCT-US95-08071-94	Sequence 94, Appl

ALIGNMENTS

RESULT 1
US-08-403-378B-3
; Sequence 3, Application US/08403378B
; Patent No. 5759991
; GENERAL INFORMATION:
; APPLICANT: TOHDOH, NAOKI
; APPLICANT: TOJO, SHIN-ICHIRO
; APPLICANT: KOJIMA, SHIN-ICHI
; APPLICANT: UEKI, YASUYUKI
; APPLICANT: NISHIHARA, TOSHIO
; APPLICANT: FUKUSHIMA, NOBUYUKI
; APPLICANT: IRIE, TSUNEMASA
; APPLICANT: AGUI, HIDEO
; APPLICANT: OJIKI, KOSEI
; TITLE OF INVENTION: NEUTROPHIC PEPTIDE DERIVATIVES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, NW
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403.378B
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-124688
; FILING DATE: 27-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-080398
; FILING DATE: 30-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-280590
; FILING DATE: 27-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-333241
; FILING DATE: 21-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-243003
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/758,043

[illegible]

SEQUENCE CHARACTERISTICS:
LENGTH: 1447 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: placental tissue
IMMEDIATE SOURCE:
CLONE: pl-3
FEATURE:
NAME/KEY: CDS
LOCATION: 120..680
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..680
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 681..1447
US-08-403-378B-14

Query Match 5.0%; Score 41; DB 2; Length 1447;
Best Local Similarity 60.2%; Pred. No. 0.0027;
Matches 68; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Qy 281 ctatatctgtgtagtgatccagatgcccttagcagacagaaacccacagacagatt 340
Db 308 CTACACCTTGGTCTGACAGACCCGGATGCTCCAGCAGGAAGATCCCAAATACAGAGA 367
Qy 341 ctgagacattgctggttaacagatatcaaggcgccgacctgaagaaaggga 393
Db 368 ATGCCATATTCTCTGTTGTCACATCAAGGGCAATGACATCAGCAGTGCSA 420

RESULT 3
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Iardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 4.2%; Score 35; DB 1; Length 7218;
Best Local Similarity 3.0%; Pred. No. 0.55;
Matches 11; Conservative 199; Mismatches 159; Indels 0; Gaps 0;
Qy 65 gatggtgctactggagacagaggatgagacagcccgctgtgcccatgagccctcctgga 124
Db 1420 RRR 1361
Qy 125 cgaggacacccctctttgccaggcgcttgaagtttaccacagagattgggggaacattgg 184
Db 1360 RRR 1301
Qy 185 ctgcaaggtgttctctgattgtacaactacagacagagatcacctcctctgtagagacc 244
Db 1300 RRR 1241
Qy 245 gatagtcgaattcccgggcgctggacgagcgcaacctatctcctggtgatggtgattcc 304
Db 1240 RRR 1181
Qy 305 agatgccccctagcagacagacccacagacagagatcttggagacattggctggttaacaga 364
Db 1180 RRR 1121
Qy 365 tatcaaggcgcccgacctgaagaaaggaagattcagggccagagattatcagccctacca 424
Db 1120 RRR 1061
Qy 425 ggcctccctc 433
Db 1060 AGCTCCCTC 1052

RESULT 4
US-07-644-372-1
Sequence 1, Application US/07644372
Patent No. 5416009
GENERAL INFORMATION:
APPLICANT: Lazzeri, Mario E.
APPLICANT: Nutman, Thomas B.
APPLICANT: Weiss, Niklaus
TITLE OF INVENTION: A DNA SEGMENT ENCODING A SPECIFIC
TITLE OF INVENTION: IMMUNODIAGNOSTIC ANTIGEN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1615 L. Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

RESULT 8
US-08-832-883-53/c
; Sequence 53, Application US/08832883
; Patent No. 5807681
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio

```

;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/832,877
;
;

```

```
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-832-877-53

Query Match          3.9%; Score 32.2; DB 3; Length 4586;
Best Local Similarity 66.7%; Pred. No. 3.2;
Matches 46; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 758 tgaacccctctggatcacagaacccctcttttccaaataaaaaaatcatccaaaa 817
Db 2505 TGCACCTCCACCTGGGTGAAGAGCGAGATCCATCTCAAAAAAAAAAAAAAAAAAAAA 2446
QY 818 aaaaaaaaa 826
Db 2445 AAAAAAAAAA 2437

RESULT 10
US-08-467-607-2
; Sequence 2, Application US/08467607
; Patent No. 5783434
; GENERAL INFORMATION:
; APPLICANT: TUNG, JAY S.
; APPLICANT: SINHA, SUKANTO
; APPLICANT: MCCONLOGUE, LISA
; APPLICANT: TATSUNO, GWEN
; APPLICANT: ANDERSON, JOHN
; APPLICANT: CHRISLER, SUSANNA
; TITLE OF INVENTION: NOVEL CATHEPSIN AND METHODS AND
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ATHENA NEUROSCIENCES
; STREET: 800 F. Gateway Blvd.
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,607
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: DUVAL, JEAN M.
; REGISTRATION NUMBER: 32,731
; REFERENCE/DOCKET NUMBER: 002010-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 877-0900
; TELEFAX: (415) 877-8370
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1558 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1011
; US-08-467-607-2

Query Match          3.9%; Score 32.2; DB 3; Length 4586;
Best Local Similarity 66.7%; Pred. No. 3.2;
Matches 46; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 758 tgaacccctctggatcacagaacccctcttttccaaataaaaaaatcatccaaaa 817
Db 2505 TGCACCTCCACCTGGGTGAAGAGCGAGATCCATCTCAAAAAAAAAAAAAAAAAAAAA 2446
QY 818 aaaaaaaaa 826
Db 2445 AAAAAAAAAA 2437

RESULT 11
US-08-469-362-2
; Sequence 2, Application US/08469362
; Patent No. 5849711
; GENERAL INFORMATION:
; APPLICANT: TUNG, JAY S.
; APPLICANT: SINHA, SUKANTO
; APPLICANT: MCCONLOGUE, LISA
; APPLICANT: SEMKO, CHRISTOPHER M.F.
; TITLE OF INVENTION: NOVEL CATHEPSIN AND METHODS AND
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ATHENA NEUROSCIENCES
; STREET: 800 F. Gateway Blvd.
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,362
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: DUVAL, JEAN M.
; REGISTRATION NUMBER: 32,731
; REFERENCE/DOCKET NUMBER: 002010-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 877-0900
; TELEFAX: (415) 877-8370
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1558 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1011
; US-08-469-362-2

Query Match          3.8%; Score 31.6; DB 2; Length 1558;
Best Local Similarity 62.5%; Pred. No. 2.7;
Matches 40; Conservative 4; Mismatches 20; Indels 0; Gaps 0;

QY 763 cccctctggatcacagaacccctcttttccaaataaaaaaatcatccaaaaaaa 822
Db 1444 CCANTGNTGRAAAARAATCTGCCCTTCGCGAAAAAAAAAAAAAAAAAAAAAAA 1503
QY 823 aaaa 826
Db 1504 AAAA 1507

STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 103..1011
US-08-467-607-2

Query Match          3.8%; Score 31.6; DB 2; Length 1558;
Best Local Similarity 62.5%; Pred. No. 2.7;
Matches 40; Conservative 4; Mismatches 20; Indels 0; Gaps 0;

QY 763 cccctctggatcacagaacccctcttttccaaataaaaaaatcatccaaaaaaa 822
Db 1444 CCANTGNTGRAAAARAATCTGCCCTTCGCGAAAAAAAAAAAAAAAAAAAAAAA 1503
QY 823 aaaa 826
Db 1504 AAAA 1507

STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 103..1011
US-08-469-362-2

Query Match          3.8%; Score 31.6; DB 3; Length 1558;
Best Local Similarity 62.5%; Pred. No. 2.7;
Matches 40; Conservative 4; Mismatches 20; Indels 0; Gaps 0;

QY 763 cccctctggatcacagaacccctcttttccaaataaaaaaatcatccaaaaaaa 822
```


RESULT 13
US-08-300-903A-8

```

RESULT 14
US-08-632-598-1/c
; Sequence 1, Application US/08632598
; Patent No. 5886164
; GENERAL INFORMATION:
; APPLICANT: BIRD, COLIN R
; APPLICANT: FLETCHER, JONATHAN D
; TITLE OF INVENTION: RIPENING-RELATED GENES FROM BANANA
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABRY AND CUSHMAN

```

```
; STREET: 1100 NEW YORK AVENUE N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/632,598
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 223355/SEE50112/US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 861-3000
; TELEFAX: 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 1712 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: MUSA
; IMMEDIATE SOURCE:
; CLONE: ACS GENE
;
US-08-632-598-1
```

```
Query Match          3.8%; Score 31.6; DB 3; Length 1712;
Best Local Similarity 65.7%; Pred. No. 2.8;
Matches 46; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
```

```
QY 757 atggaaacccctctggtacagaccctctttccaaataaaaaaatcatccaaa 816
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1614 ATATACTTAGTGGACATCAACCCATTTTCCCAAAAAAATAAATATCAAA 1555

QY 817 aaaaaaaa 826
|| ||| ||
Db 1554 GAACAAAGAA 1545
```

```
RESULT 15
US-08-738-349-1
; Sequence 1, Application US/08738349
; Patent No. 5869638
; GENERAL INFORMATION:
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Kawai, Shinji
; APPLICANT: Tsujimura, Atsushi
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Cadherin-Like Protein and
; TITLE OF INVENTION: Process for Its Production
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,349
; FILING DATE: 25-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/364,439
; FILING DATE:
; FILING DATE: 26-AUG-1993
; REFERENCE/DOCKET NUMBER: 02481.1323-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mrna
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; STRAIN: osteoblastic cell line MC3T3E1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 284..2671
;
US-08-738-349-1
```

```
Query Match          3.8%; Score 31.6; DB 3; Length 3581;
Best Local Similarity 74.1%; Pred. No. 4.3;
Matches 40; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 773 atacagaaacccctctttccaaataaaaaaatcatccaaaaa 826
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3483 ATAAAAAAGATTCCTTTTAAAAA 3536
```

```
Search completed: May 23, 2000, 23:56:59
Job time: 10531 sec
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2000, 01:02:48 ; Search time 3550.87 Seconds
(without alignments)
-969.808 Million cell updates/sec

Title: US-09-215-435-124
Perfect score: 826
Sequence: 1 aaccagagtgccatgggt.....atcatcctcaaaaaaaaaa 826

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5099991 seqs, -2084540658 residues

Total number of hits satisfying chosen parameters: 10199964

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA:*

1: /cgn2_6/ptodata/1/pna/US092C_COMB.seq:*

2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/1/pna/US081A_COMB.seq:*

6: /cgn2_6/ptodata/1/pna/US081B_COMB.seq:*

7: /cgn2_6/ptodata/1/pna/US081C_COMB.seq:*

8: /cgn2_6/ptodata/1/pna/US081D_COMB.seq:*

9: /cgn2_6/ptodata/1/pna/US082A_COMB.seq:*

10: /cgn2_6/ptodata/1/pna/US082B_COMB.seq:*

11: /cgn2_6/ptodata/1/pna/US082C_COMB.seq:*

12: /cgn2_6/ptodata/1/pna/US083A_COMB.seq:*

13: /cgn2_6/ptodata/1/pna/US083B_COMB.seq:*

14: /cgn2_6/ptodata/1/pna/US084A_COMB.seq:*

15: /cgn2_6/ptodata/1/pna/US084B_COMB.seq:*

16: /cgn2_6/ptodata/1/pna/US084C_COMB.seq:*

17: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*

18: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*

19: /cgn2_6/ptodata/1/pna/US087A_COMB.seq:*

20: /cgn2_6/ptodata/1/pna/US087B_COMB.seq:*

21: /cgn2_6/ptodata/1/pna/US087C_COMB.seq:*

22: /cgn2_6/ptodata/1/pna/US088A_COMB.seq:*

23: /cgn2_6/ptodata/1/pna/US088B_COMB.seq:*

24: /cgn2_6/ptodata/1/pna/US088C_COMB.seq:*

25: /cgn2_6/ptodata/1/pna/US088D_COMB.seq:*

26: /cgn2_6/ptodata/1/pna/US089A_COMB.seq:*

27: /cgn2_6/ptodata/1/pna/US089B_COMB.seq:*

28: /cgn2_6/ptodata/1/pna/US089C_COMB.seq:*

29: /cgn2_6/ptodata/1/pna/US089D_COMB.seq:*

30: /cgn2_6/ptodata/1/pna/US089E_COMB.seq:*

31: /cgn2_6/ptodata/1/pna/US089F_COMB.seq:*

32: /cgn2_6/ptodata/1/pna/US089G_COMB.seq:*

33: /cgn2_6/ptodata/1/pna/US089H_COMB.seq:*

34: /cgn2_6/ptodata/1/pna/US089I_COMB.seq:*

35: /cgn2_6/ptodata/1/pna/US090A_COMB.seq:*

36: /cgn2_6/ptodata/1/pna/US090B_COMB.seq:*

37: /cgn2_6/ptodata/1/pna/US090C_COMB.seq:*

38: /cgn2_6/ptodata/1/pna/US090D_COMB.seq:*

39: /cgn2_6/ptodata/1/pna/US090E_COMB.seq:*

40: /cgn2_6/ptodata/1/pna/US091A_COMB.seq:*

41: /cgn2_6/ptodata/1/pna/US091B_COMB.seq:*

42: /cgn2_6/ptodata/1/pna/US091C_COMB.seq:*

43: /cgn2_6/ptodata/1/pna/US092A_COMB.seq:*

44: /cgn2_6/ptodata/1/pna/US092B_COMB.seq:*

45: /cgn2_6/ptodata/1/pna/US092C_COMB.seq:*

46: /cgn2_6/ptodata/1/pna/US092D_COMB.seq:*

47: /cgn2_6/ptodata/1/pna/US092E_COMB.seq:*

48: /cgn2_6/ptodata/1/pna/US093A_COMB.seq:*

49: /cgn2_6/ptodata/1/pna/US093B_COMB.seq:*

50: /cgn2_6/ptodata/1/pna/US093C_COMB.seq:*

51: /cgn2_6/ptodata/1/pna/US093D_COMB.seq:*

52: /cgn2_6/ptodata/1/pna/US093E_COMB.seq:*

53: /cgn2_6/ptodata/1/pna/US093F_COMB.seq:*

54: /cgn2_6/ptodata/1/pna/US094A_COMB.seq:*

55: /cgn2_6/ptodata/1/pna/US094B_COMB.seq:*

56: /cgn2_6/ptodata/1/pna/US094C_COMB.seq:*

57: /cgn2_6/ptodata/1/pna/US095_COMB.seq:*

58: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*

59: /cgn2_6/ptodata/1/pna/US6001A_COMB.seq:*

60: /cgn2_6/ptodata/1/pna/US6001B_COMB.seq:*

61: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq:*

62: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq:*

63: /cgn2_6/ptodata/1/pna/US6002C_COMB.seq:*

64: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq:*

65: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq:*

66: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq:*

67: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq:*

68: /cgn2_6/ptodata/1/pna/US6004C_COMB.seq:*

69: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*

70: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*

71: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*

72: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq:*

73: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq:*

74: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq:*

75: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq:*

76: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq:*

77: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq:*

78: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq:*

79: /cgn2_6/ptodata/1/pna/US6011A_COMB.seq:*

80: /cgn2_6/ptodata/1/pna/US6011B_COMB.seq:*

81: /cgn2_6/ptodata/1/pna/US6012A_COMB.seq:*

82: /cgn2_6/ptodata/1/pna/US6012B_COMB.seq:*

83: /cgn2_6/ptodata/1/pna/US6013A_COMB.seq:*

84: /cgn2_6/ptodata/1/pna/US6013B_COMB.seq:*

85: /cgn2_6/ptodata/1/pna/US6014A_COMB.seq:*

86: /cgn2_6/ptodata/1/pna/US6014B_COMB.seq:*

87: /cgn2_6/ptodata/1/pna/US6014C_COMB.seq:*

88: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*

89: /cgn2_6/ptodata/1/pna/US6016A_COMB.seq:*

90: /cgn2_6/ptodata/1/pna/US6016B_COMB.seq:*

91: /cgn2_6/ptodata/1/pna/US6016C_COMB.seq:*

92: /cgn2_6/ptodata/1/pna/US6017A_COMB.seq:*

93: /cgn2_6/ptodata/1/pna/US6017B_COMB.seq:*

94: /cgn2_6/ptodata/1/pna/US6017C_COMB.seq:*

95: /cgn2_6/ptodata/1/pna/US6018A_COMB.seq:*

96: /cgn2_6/ptodata/1/pna/US6018B_COMB.seq:*

97: /cgn2_6/ptodata/1/pna/US6018C_COMB.seq:*

98: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*

99: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*

100: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*

101: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*

102: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*

103: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*

104: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	826	100.0	826	43	US-09-215-435-124
2	826	100.0	826	56	US-09-471-276-7
3	826	100.0	826	72	US-60-081-563-70

Sequence 124, Appl
Sequence 7, Appli
Sequence 70, Appl

[illegible]

Qy 781 ccccttctttccaaataaaaaaaatcatcccaaaaaaa 826
|||||
Db 781 ccccttctttccaaataaaaaaaatcatcccaaaaaaa 826
|||||

RESULT 2
US-09-471-276-7
; Sequence 7, Application US/09471276
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: GENSET 025Cp1
; CURRENT APPLICATION NUMBER: US/09/471,276
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1622
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLGLMMVVTG/DE
US-09-471-276-7

Query Match 100.08; Score 826; DB 56; Length 826;
Best Local Similarity 100.08; Pred. No. 1.3e-169;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaccagagtgccctgggttgacaatgagctggtcacagcagcactgttactgggtc 60
|||||
Db 1 aaccagagtgccctgggttgacaatgagctggtcacagcagcactgttactgggtc 60
|||||
Qy 61 tcatgatggtgctcactgagacgagagtgagaaacagcccggtgcccagagggccctcc 120
|||||
Db 61 tcatgatggtgctcactgagacgagagtgagaaacagcccggtgcccagagggccctcc 120
|||||
Qy 121 tggacagagacaccctcttttccaggggaccttgaaagtttctaccagaggttggggaaca 180
|||||
Db 121 tggacagagacaccctcttttccaggggaccttgaaagtttctaccagaggttggggaaca 180
|||||
Qy 181 ttggctgcaagtggttctctgattgtaacaactacagacagaaagatcacctcctgagatg 240
|||||
Db 181 ttggctgcaagtggttctctgattgtaacaactacagacagaaagatcacctcctgagatg 240
|||||
Qy 241 agccgatagtcagttcccgggggccgtgacggcgcaacctatctcgtggtgagtggtg 300
|||||
Db 241 agccgatagtcagttcccgggggccgtgacggcgcaacctatctcgtggtgagtggtg 300
|||||
Qy 301 atccagatgcccttagcagacagacaccacagacagatctctggagacatttggtggttaa 360
|||||
Db 301 atccagatgcccttagcagacagacaccacagacagatctctggagacatttggtggttaa 360
|||||
Qy 361 cagatacaaggcgccgacctgaagaaggaagattcaggccagagattatcagcct 420
|||||
Db 361 cagatacaaggcgccgacctgaagaaggaagattcaggccagagattatcagcct 420
|||||
Qy 421 accaggtccctcccccagcgacacagtggttccatcgctaccagttcttggctatc 480
|||||
Db 421 accaggtccctcccccagcgacacagtggttccatcgctaccagttcttggctatc 480
|||||

Qy 481 ttccaggaaggaaggtcatctctctccctcccaaggaaaaaaactcagagctcttga 540
|||||
Db 481 ttccaggaaggaaggtcatctctctccctcccaaggaaaaaaactcagagctcttga 540
|||||
Qy 541 aaatggacagatttctgaaccgtttccacctggggcgaacctgaagcaagcaccagttca 600
|||||
Db 541 aaatggacagatttctgaaccgtttccacctggggcgaacctgaagcaagcaccagttca 600
|||||
Qy 601 tgaccagaactaccagagactcacaaacctccagagctcccgagagaaagggccagcgagc 660
|||||
Db 601 tgaccagaactaccagagactcacaaacctccagagctcccgagagaaagggccagcgagc 660
|||||
Qy 661 ccaagcacaaaaacagcgcgagatagctgctctagatagccggttgcctcccg 720
|||||
Db 661 ccaagcacaaaaacagcgcgagatagctgctctagatagccggttgcctcccg 720
|||||
Qy 721 catgtggccacactgcccaccacgacgatgtgggtatggaacccctctgatacagaa 780
|||||
Db 721 catgtggccacactgcccaccacgacgatgtgggtatggaacccctctgatacagaa 780
|||||
Qy 781 ccccttctttccaaataaaaaaaatcatcccaaaaaaa 826
|||||
Db 781 ccccttctttccaaataaaaaaaatcatcccaaaaaaa 826
|||||

RESULT 3
US-60-081-563-70
; Sequence 70, Application US/60081563
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/081,563
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.027PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Prostate
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80

```
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 8.5
OTHER INFORMATION: seq AALLGLMMVVTG/DE
FEATURE:
NAME/KEY: poly_a_signal
LOCATION: 795..800
FEATURE:
NAME/KEY: poly_a
LOCATION: 814..826
IDENTIFICATION METHOD: blastn
FEATURE:
NAME/KEY: est
LOCATION: 426..664
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 99..337
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 327..435
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 1..109
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 716..777
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 391..452
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 660..720
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 334..394
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 506..630
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 90
OTHER INFORMATION: region 327..451
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 323..435
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 94
OTHER INFORMATION: region 143..255
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 180..280
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 2..102
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 464..528
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 284..348
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 279..321
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 100..142
OTHER INFORMATION: id W32197
```

```
FEATURE:
NAME/KEY: est
LOCATION: 426..466
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 245..285
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 358..674
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 141..457
OTHER INFORMATION: id AAL192427
FEATURE:
NAME/KEY: est
LOCATION: 670..779
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 37..146
OTHER INFORMATION: id AAL192427
FEATURE:
NAME/KEY: est
LOCATION: 485..602
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 224..341
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 601..709
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 91
OTHER INFORMATION: region 115..223
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 704..779
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 46..121
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 392..436
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 393..437
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 429..465
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 94
OTHER INFORMATION: region 363..399
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 464..493
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 334..363
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 343..730
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 8..395
OTHER INFORMATION: id HSPD04880
FEATURE:
NAME/KEY: est
LOCATION: 507..663
```


IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 96
; OTHER INFORMATION: region 264..420
; OTHER INFORMATION: id AA192426
US-60-081-563-70

Query Match 100.0%; Score 826; DB 72; Length 826;
Best Local Similarity 100.0%; Pred. No. 1.3e-169; Indels 0; Gaps 0;
Matches 826; Conservative 0; Mismatches 0

```
Qy 1 aaccagagtgcccatgggttggaacatgagctgggtcacagcagcactgttactgggtc 60
Db 1 AACCAGAGTGCCCATGGGTGGCAATGAGCTGGTCACAGCAGCACTGTTACTGGGTC 60

Qy 61 tcatgatggtgctcactgagagcagagatgagaacagcccggtgccccatgagccctcc 120
Db 61 TCATGATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCATGAGGCCCTCC 120

Qy 121 tggcagagacacccctctttgcccagggccttgaagttttaccacagagttggggaaca 180
Db 121 TGGCAGGACACCCCTCTTTGCCAGGGCCCTGAAGTTTCTACCCAGAGTTGGGAACA 180

Qy 181 ttggctgcaagtggttctctgattgttaacaaactacagacagaagatcacctcctggatgg 240
Db 181 TTGGCTGCAAGTGTCTCTGATTGTAACTACAGACAGAAAGATCACCTCTGGATGG 240

Qy 241 agccgatagcaagtcccgggggccgtggacgcgcgaacacctatctctgtgatgggtgg 300
Db 241 AGCCGATAGTCAAGTTCGCCGGGGCCGTGGACGCGCAACCTATATCTGGTGTGATGGTGG 300

Qy 301 atccagatgcccctagacagacaaacccagacagagattctgagacattggctgggttaa 360
Db 301 ATCCAGATGCCCTTAGCAGACGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAA 360

Qy 361 cagatatcaaggcgccgacacctgaagaaggaaagattcaggggccagaggttatcagcct 420
Db 361 CAGATATCAAGGGCCGACCTGAAGAAAGGAAAGATTCAAGGCCAGGAGTATCAGCCT 420

Qy 421 accaggtccctccccaccggcacacagtggtcttcacgtaccagttcttctgtctatc 480
Db 421 ACCAGGTCCTCCCTCCACCGGCACACAGTGGCTTCCATCGCTACCCAGTCTCTTGTCTATC 480

Qy 481 ttcaggaagaaagtgcatctctctcccaaggaaacaaactcgagctcttggaa 540
Db 481 TTCAGGAAGAAAGTGTCATCTCTCTCTCCCAAGGAAACAAACTCGAGGCTCTTGGAA 540

Qy 541 aaatggacagatttctgaacctttccacctgggcgaacctgaagcaagcaccagttca 600
Db 541 AAATGGACAGATTCTGAACCGTTTCCACCTGGCGGAACCTGAAGCAAGCACCCAGTTCA 600

Qy 601 tgaccagaaactaccagagactcaacacccctccaggtctccagagaaagggccagcgagc 660
Db 601 TGACCCAGAACTACGAGGACTCACCAACCCCTCCAGGCTCCAGAGAAAGGGCCAGCGAGC 660

Qy 720 ccaagcacaaaaaccagcgagatagctgctgtagatagcgggttggccatccggg 720
Db 720 CCAAGCACAAAAACGAGCGGAGATAGCTGCTGCTAGATAGCGGGCTTTCATCTCCGCGG 720

Qy 781 catgtggccacactgcccaccacagcagatgtgggtatggaaacccctctggtatcacgaa 780
Db 781 CATGTGGCCACACTGCCACACCGACGATGTGGGTATGGAAACCCCTCTGTGATACAGAA 780

Qy 826 ccccttctttccaaataaaaaaaatcatccaaaaaa 826
Db 826 CCCCTTCTTTTCCAAATAAAAAAATCATCCAAAAA 826
```

RESULT 4
US-60-147-499-7
; Sequence 7, Application US/60147499
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.

APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/60/147.499
; CURRENT FILING DATE: 1999-08-05
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLGLMMVVTG/DE
US-60-147-499-7

Query Match 100.0%; Score 826; DB 87; Length 826;
Best Local Similarity 100.0%; Pred. No. 1.3e-169;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 aaccagagtgcccatgggttggaacatgagctgggtcacagcagcactgttactgggtc 60
Db 1 aaccagagtgcccatgggttggaacatgagctgggtcacagcagcactgttactgggtc 60

Qy 61 tcatgatggtgctcactgagacagagatgagaacagcccggtgccccatgagccctcc 120
Db 61 tcatgatggtgctcactgagacagagatgagaacagcccggtgccccatgagccctcc 120

Qy 121 tggcagagacacccctctttgcccagggccttgaagttttaccacagagttggggaaca 180
Db 121 tggcagagacacccctctttgcccagggccttgaagttttaccacagagttggggaaca 180

Qy 181 ttggctgcaagtggttctctgattgttaacaaactacagacagaagatcacctcctggatgg 240
Db 181 ttggctgcaagtggttctctgattgttaacaaactacagacagaagatcacctcctggatgg 240

Qy 241 agccgatagcaagtcccgggggccgtggacgcgcgaacacctatctctgtgatgggtgg 300
Db 241 agccgatagcaagtcccgggggccgtggacgcgcgaacacctatctctgtgatgggtgg 300

Qy 301 atccagatgcccctagacagacaaacccagacagagattctgagacattggctgggttaa 360
Db 301 atccagatgcccctagacagacaaacccagacagagattctgagacattggctgggttaa 360

Qy 361 cagatatcaaggcgccgacacctgaagaaggaaagattcaggggccagaggttatcagcct 420
Db 361 cagatatcaaggcgccgacacctgaagaaggaaagattcaggggccagaggttatcagcct 420

Qy 421 accaggtccctccccaccggcacacagtggtcttcacgtaccagttcttctgtctatc 480
Db 421 accaggtccctccccaccggcacacagtggtcttcacgtaccagttcttctgtctatc 480

Qy 481 ttcaggaagaaagtgcatctctctcccaaggaaacaaactcgagctcttggaa 540
Db 481 ttcaggaagaaagtgcatctctctcccaaggaaacaaactcgagctcttggaa 540

Qy 541 aaatggacagatttctgaacctttccacctgggcgaacctgaagcaagcaccagttca 600
Db 541 aaatggacagatttctgaacctttccacctgggcgaacctgaagcaagcaccagttca 600

Qy 601 tgaccagaaactaccagagactcaacacccctccaggtctccagagaaagggccagcgagc 660
Db 601 tgaccagaaactaccagagactcaacacccctccaggtctccagagaaagggccagcgagc 660

Qy 661 ccaagcacaaaaaccagcgagatagctgctgtagatagcgggttggccatccggg 720
Db 661 ccaagcacaaaaaccagcgagatagctgctgtagatagcgggttggccatccggg 720
```

```
|||||
Db 661 ccaagcaaaaaaacaggcggagatagctgctgctagatagcgggtttgccaatccggg 720
Qy 721 catgtggccacactgccaccaccagcagatgtgggtatgaacccctctggatcacagaa 780
Db 721 catgtggccacactgccaccaccagcagatgtgggtatgaacccctctggatcacagaa 780
Qy 781 cccctttttccaaataaaaaaaatcatccaaaaaaataaaaaa 826
Db 781 cccctttttccaaataaaaaaaatcatccaaaaaaataaaaaa 826

RESULT 5
US-60-169-629-7
; Sequence 7, Application US/60169629
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Bougueleret, L.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: cDNAs for Secreted Proteins
; FILE REFERENCE: GENSET.071PRF
; CURRENT APPLICATION NUMBER: US/60/169,629
; CURRENT FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 715
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLGLMVMVVG/DE
US-60-169-629-7

Query Match 100.0%; Score 826; DB 91; Length 826;
Best Local Similarity 100.0%; Pred. No. 1.3e-169;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaccagaggtgccatggttgacaatgagctggtcacagcagcactgttactgggtc 60
Db 1 aaccagaggtgccatggttgacaatgagctggtcacagcagcactgttactgggtc 60
Qy 61 tcatgatggttggttcaactggagacgagatgagaacagccgtgtgccatgagccctcc 120
Db 61 tcatgatggttggttcaactggagacgagatgagaacagccgtgtgccatgagccctcc 120
Qy 121 tggacgagagacacccttttgcaggcccttgaagttttctaccacagagttatcagcct 180
Db 121 tggacgagagacacccttttgcaggcccttgaagttttctaccacagagttatcagcct 180
Qy 181 ttggctgaaggttttcctgattgtaacaactacagacagaagatcacctctggatgg 240
Db 181 ttggctgaaggttttcctgattgtaacaactacagacagaagatcacctctggatgg 240
Qy 241 agccgatagtcgaagttccccggggccgtggacgcccgaacctatctctggtgatggtg 300
Db 241 agccgatagtcgaagttccccggggccgtggacgcccgaacctatctctggtgatggtg 300
Qy 301 atccagatgcccttagcagacgaacccacagacagattctgagacattctggtggttaa 360
Db 301 atccagatgcccttagcagacgaacccacagacagattctggtggtggttaa 360
Qy 361 cagatatcaaggccgcagcctgaagaaagggaagattcaggccagggaggttatcagcct 420
Db 361 cagatatcaaggccgcagcctgaagaaagggaagattcaggccagggaggttatcagcct 420
```

```
Qy 421 accaggtccctcccccacccggcgcacacagtggttccatcgctaccagttctttgtctatc 480
Db 421 accaggtccctcccccacccggcgcacacagtggttccatcgctaccagttctttgtctatc 480
Qy 481 ttacgaagaagaaggttcattctctctctcccaaggagaaacaaactcgaggctcttggga 540
Db 481 ttacgaagaagaaggttcattctctctctcccaaggagaaacaaactcgaggctcttggga 540
Qy 541 aaatggacagatttctgaacctttccacctgggcggaacctgaagcagcaccctcagttca 600
Db 541 aaatggacagatttctgaacctttccacctgggcggaacctgaagcagcaccctcagttca 600
Qy 601 tgaccagaactaccaggaactcaccacccctccaggtccacagagaagggccagcgagc 660
Db 601 tgaccagaactaccaggaactcaccacccctccaggtccacagagaagggccagcgagc 660
Qy 661 ccaagcacaaaaaccaggcggagatagctcctgctagatagccggtttgccaatccggg 720
Db 661 ccaagcacaaaaaccaggcggagatagctcctgctagatagccggtttgccaatccggg 720
Qy 721 catgtggccacactgccaccaccagcagatgtgggtatgaacccctctggatcacagaa 780
Db 721 catgtggccacactgccaccaccagcagatgtgggtatgaacccctctggatcacagaa 780
Qy 781 cccctttttccaaataaaaaaaatcatccaaaaaaataaaaaa 826
Db 781 cccctttttccaaataaaaaaaatcatccaaaaaaataaaaaa 826

RESULT 6
US-60-187-470-7
; Sequence 7, Application US/60187470
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Bougueleret, L.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: cDNAs for Secreted Proteins
; FILE REFERENCE: 78.US2.PRO
; CURRENT APPLICATION NUMBER: US/60/187,470
; CURRENT FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 715
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLGLMVMVVG/DE
US-60-187-470-7

Query Match 100.0%; Score 826; DB 97; Length 826;
Best Local Similarity 100.0%; Pred. No. 1.3e-169;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaccagaggtgccatggttgacaatgagctggtcacagcagcactgttactgggtc 60
Db 1 aaccagaggtgccatggttgacaatgagctggtcacagcagcactgttactgggtc 60
Qy 61 tcatgatggttggttcaactggagacgagatgagaacagccgtgtgccatgagccctcc 120
Db 61 tcatgatggttggttcaactggagacgagatgagaacagccgtgtgccatgagccctcc 120
Qy 121 tggacgagagacacccttttgcaggcccttgaagttttctaccacagagttatgggaaca 180
Db 121 tggacgagagacacccttttgcaggcccttgaagttttctaccacagagttatgggaaca 180
```


US-09-359-922-5157
; Sequence 5157, Application US/09359922
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 5157
; LENGTH: 878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-5157

Query Match 98.2%; Score 810.8; DB 51; Length 878;
Best Local Similarity 99.1%; Pred. No. 2.7e-166; Mismatches 0; Indels 0; Gaps 0;
Matches 815; Conservative 0;

QY 1 aaccagaggtgccatgggttgacaatgaggtggtcacagcagcactgttactgggtc 60
Db 57 aaccagaggtgccatgggttgacaatgaggtggtcacagcagcactgttactgggtc 116
QY 61 tcattgatgttggttcactggagacagaggtgagaaaccccggtgcccattgagccctcc 120
Db 117 tcattgatgttggttcactggagacagaggtgagaaaccccggtgcccattgagccctct 176
QY 121 tggacgagacacccctctttgcccagggcctgaagtttttccaccagagttgggaaca 180
Db 177 tggacgagacacccctctttgcccagggcctgaagtttttccaccagagttgggaaca 236
QY 181 ttggctgcaaggtgttcctgattgtaacaactacagacagagattcctcctggtgag 240
Db 237 ttggctgcaaggtgttcctgattgtaacaactacagacagagattcctcctggtgag 296
QY 241 agccgatagtcgaagttcccgccggtgagccggtgagccggtgagccggtgagccggt 300
Db 297 agccgatagtcgaagttcccgccggtgagccggtgagccggtgagccggtgagccggt 356
QY 301 atccagatgccctagcagacagacccagacagagattcctgagacattggtggttaa 360
Db 357 atccagatgccctagcagacagacccagacagagattcctgagacattggtggttaa 416
QY 361 cagatatcaaggcgccgacctgaagaagggaagattccaggccaggagttatcagcct 420
Db 417 cagatatcaaggcgccgacctgaagaagggaagattccaggccaggagttatcagcct 476
QY 421 accaggtccctcccccagccagccagtcacagtggttccatcgctaccggtttttgttctatc 480
Db 477 accaggtccctcccccagccagccagtcacagtggttccatcgctaccggtttttgttctatc 536
QY 481 ttccaggaagaaaggttcattctctctctcccaagaaaggaagattccaggccaggagttcagcct 540
Db 537 ttccaggaagaaaggttcattctctctctcccaagaaaggaagattccaggccaggagttcagcct 596
QY 541 aataggacagatttctgaacccgtttccacctggcgcaacctgaagcaagcaccagttca 600
Db 597 aataggacagatttctgaacccgtttccacctggcgcaacctgaagcaagcaccagttca 656
QY 601 tgaccagaaactaccaggaactccaaacctccaggtcccccagagaagggccagcgagc 660
Db 657 tgaccagaaactaccaggaactccaaacctccaggtcccccagagaagggccagcgagc 716
QY 661 ccaagcacaaaaaccagcgagatagctgcctgcttagatagccggtttgcccattccggg 720
Db 717 ccaagcacaaaaaccagcgagatagctgcctgcttagatagccggtttgcccattccggg 776
QY 721 catgtggccacactgccaccaccagcagatgtgggtatggaaacccctctggatacagaa 780

Db 777 catgtggccacactgccaccaccgacgatgtgggtatggaaacccctctctggatacagaa 836
QY 781 cccctctctttccaaataaaaaaaatcatccaaaaa 822
Db 837 cccctctctttccaaataaaaaaaatcatccagga 878

RESULT 9
US-09-359-922-5157
; Sequence 5157, Application US/09359922A
; GENERAL INFORMATION:
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Liu, Jin
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922A
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: US 09/034,341
; EARLIER FILING DATE: 1998-02-13
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 5157
; LENGTH: 878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-5157

Query Match 98.2%; Score 810.8; DB 51; Length 878;
Best Local Similarity 99.1%; Pred. No. 2.7e-166; Mismatches 0; Indels 0; Gaps 0;
Matches 815; Conservative 0;

QY 1 aaccagaggtgccatgggttgacaatgaggtggtcacagcagcactgttactgggtc 60
Db 57 aaccagaggtgccatgggttgacaatgaggtggtcacagcagcactgttactgggtc 116
QY 61 tcattgatgttggttcactggagacagaggtgagaaaccccggtgcccattgagccctcc 120
Db 117 tcattgatgttggttcactggagacagaggtgagaaaccccggtgcccattgagccctct 176
QY 121 tggacgagacacccctctttgcccagggcctgaagtttttccaccagagttgggaaca 180
Db 177 tggacgagacacccctctttgcccagggcctgaagtttttccaccagagttgggaaca 236
QY 181 ttggctgcaaggtgttcctgattgtaacaactacagacagagattcctcctggtgag 240
Db 237 ttggctgcaaggtgttcctgattgtaacaactacagacagagattcctcctggtgag 296
QY 241 agccgatagtcgaagttcccgccggtgagccggtgagccggtgagccggtgagccggt 300
Db 297 agccgatagtcgaagttcccgccggtgagccggtgagccggtgagccggtgagccggt 356
QY 301 atccagatgccctagcagacagacccagacagagattcctgagacattggtggttaa 360
Db 357 atccagatgccctagcagacagacccagacagagattcctgagacattggtggttaa 416
QY 361 cagatatcaaggcgccgacctgaagaagggaagattccaggccaggagttatcagcct 420
Db 417 cagatatcaaggcgccgacctgaagaagggaagattccaggccaggagttatcagcct 476
QY 421 accaggtccctcccccagccagccagtcacagtggttccatcgctaccggtttttgttctatc 480
Db 477 accaggtccctcccccagccagccagtcacagtggttccatcgctaccggtttttgttctatc 536
QY 481 ttccaggaagaaaggttcattctctctctcccaagaaaggaagattccaggccaggagttcagcct 540
Db 537 ttccaggaagaaaggttcattctctctctcccaagaaaggaagattccaggccaggagttcagcct 596
QY 541 aataggacagatttctgaacccgtttccacctggcgcaacctgaagcaagcaccagttca 600

```
|||||
Db 597 aaatggacagatttctgaacgcgttccacgttccacgttggcggaacctgaagcaagcaccagttca 656
Qy 601 tgacccagaactaccaggaactaccacacccctccaggtccacagagaagaagggccagcgagc 660
Db 657 tgacccagaactaccaggaactaccacacccctccaggtccacagagaagaagggccagcgagc 716
Qy 661 ccaagcacaaaaaacccagcgagagatagctgcctgctagatagcgggttttgccatccggg 720
Db 717 ccaagcacaaaaaacccagcgagagatagctgcctgctagatagcgggttttgccatccggg 776
Qy 721 catgtggccacactgccacacacacgacgatgtgggtatggaacccccctctggatacagaa 780
Db 777 catgtggccacactgccacacacacgacgatgtgggtatggaacccccctctggatacagaa 836
Qy 781 ccccttttttccaaataaaaaaaatcatccaaaaaaa 822
Db 837 ccccttttttccaaataaaaaaaatcatccaggaaaaaa 878

RESULT 10
; Sequence 10894, Application US/60172373
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymorph
; FILE REFERENCE: GX-0006 P
; CURRENT APPLICATION NUMBER: US/60/172,373
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 25,772
; SOFTWARE: PERL Program
; SEQ ID NO 10894
; LENGTH: 904
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 204392.1
US-60-172-373-10894

Query Match 97.8% Score 807.6; DB 92; Length 904;
Best Local Similarity 99.5%; Pred. No. 1.3e-165;
Matches 810; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aaccagaggtgccatgggttggaacaatgagctgggtccacagcagcactgttactgggtc 60
Db 75 aaccagaggtgccatgggttggaacaatgagctgggtccacagcagcactgttactgggtc 134
Qy 61 tcattgatgtgttggttcactggagacagagatgagaacaccccggtgccccatgagggccctcc 120
Db 135 tcattgatgtgttggttcactggagacagagatgagaacaccccggtgccccatgagggccctc 194
Qy 121 tggagaggagacacccctttttgcccagggccttgaagtttttaccacagagttggggaaca 180
Db 195 tggagaggagacacccctttttgcccagggccttgaagtttttaccacagagttggggaaca 254
Qy 181 ttggtcgaaggtgttcctgattgttaacactacagacagaagatcacctccctggatgg 240
Db 255 ttggtcgaaggtgttcctgattgttaacactacagacagaagatcacctccctggatgg 314
Qy 241 agccgatagtcgaattcccccggggccgtggacggcgcaacctatatactcgtgatgggtgg 300
Db 315 agccgatagtcgaattcccccggggccgtggacggcgcaacctatatactcgtgatgggtgg 374
Qy 301 atccagatgccccttagacagacagaacacccacagacagagattctgagacattggctgtaa 360
Db 375 atccagatgccccttagacagacagaacacccacagacagagattctgagacattggctgtaa 434
Qy 361 cagatatcaaggcgccgacacctgaagaaagggaagattcaggggccaggagttatcagcct 420
```

```
|||||
Db 435 cagatatcaaggcgccgacacctgaagaaagggaagattcaggggccaggagttatcagcct 494
Qy 421 accaggtccctcccccaccccggcacacagtggtctccatcgctaccagttctttgtctatc 480
Db 495 aceaggtccctcccccaccccggcacacagtggtctccatcgctaccagttctttgtctatc 554
Qy 481 ttcagggaagaaaggtcatctctctcccaaggaacaaacaaacaaactcgaggtcttgga 540
Db 555 ttcagggaagaaaggtcatctctctcccaaggaacaaacaaacaaactcgaggtcttgga 614
Qy 541 aaatggacagatttctgaacccgtttccacccctggcggaacctgaagcaagcaccagttca 600
Db 615 aaatggacagatttctgaacccgtttccacccctggcggaacctgaagcaagcaccagttca 674
Qy 601 tgacccagaactaccaggaactaccacacccctccaggtccacagagaagaagggccagcgagc 660
Db 675 tgacccagaactaccaggaactaccacacccctccaggtccacagagaagaagggccagcgagc 734
Qy 661 ccaagcacaaaaaacccagcgagagatagctgcctgctagatagcgggtttgcatccggg 720
Db 735 ccaagcacaaaaaacccagcgagagatagctgcctgctagatagcgggtttgcatccggg 794
Qy 721 catgtggccacactgccacacacacgacgatgtgggtatggaacccccctctggtacacagaa 780
Db 795 catgtggccacactgccacacacacgacgatgtgggtatggaacccccctctggtacacagaa 854
Qy 781 ccccttttttccaaataaaaaaaatcatcca 814
Db 855 ccccttttttccaaataaaaaaaatcatcca 888

RESULT 11
US-08-958-820-4
; Sequence 4, Application US/08958820
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/958,820
; FILING DATE: Filed Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0379 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGTUT12
CLONE: 3126479
US-08-958-820-4

Query Match 97.6%; Score 806; DB 30; Length 903;
Best Local Similarity 99.4%; Pred. No. 3e-165;
Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 aaccagaggtgccatgggttgacaatgagctggtcacagcagcactgttactgggtc 60
DB 74 AACACAGAGGTGCCATGGGTGGACATGAGGCTGTGTACAGCAGCAGCTGTTACTGGGTC 133
QY 61 tcatgatgtgttcaactgagcagagagatgagaacagccgtgtgcccattgagccctcc 120
DB 134 TCATGATGTGGTCACTGGAGAGAGAGATGAGAACAGCCGCTGTGCCATGAGGCCCTCT 193
QY 121 tggacagagacacccctctttgcccagggccttgaagtcttaccagagagttggggaaca 180
DB 194 TGGACAGGAGACACCCCTCTTTTGGCAGGGCCTTCAAGTTTTCTACCCAGAGATTGGGGAACA 253
QY 181 ttggttgcgaaggtgttctctgattgtaaacactacagacagaagatcacctcctgagtg 240
DB 254 TTGGCTGCAAGGTGTTCCTGATTGTAACAACTACAGACAGAAAGATCACCTCCTGGATGG 313
QY 241 agccgatagtcgaagttcccggtgggcccgtgtgacgcgcgaacctatactcctggtgagtg 300
DB 314 AGCCGATAGTCAAGTTCCTGGGGGCGCTGAGCGGCGCAACCTATATCTTGGTATGGTGG 373
QY 301 atccagatgccccctagcagagcagaacccagacagagattctgagacattggtggtgtaa 360
DB 374 ATCCAGATGCCCCTAGCAGAGCAGAACCCACAGACAGAGATTCTGGAGACATTGGCTGGTAA 433
QY 361 cagatatcaagggtgcgcactgaagaaagggaagattccagggtccaggaggtatcagcct 420
DB 434 CAGATATCAAGGGCGCGACCTGAAGGAAGGAAGATTTCAGGGCCAGGAGTTATCAGCCT 493
QY 421 accaggtccctcccaaccggccacacagtggttccatcgctaccagttcttctgtatc 480
DB 494 ACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAAGTTCTTTGTCTATC 553
QY 481 ttcagggaagaaaggtcatctctctctctcccaaggaaacaaactcagaggtccttggga 540
DB 554 TTCAGGAAGGAAGTCACTCTCTCTCTCCCAAGGAAGAAACAAACTCGAGGCTCTTGA 613
QY 541 aaatggacagatttctgaaccgtttccacactgggcgaacctgaagcaagcaccagttca 600
DB 614 AAATGGACAGATTCTGAACCGCTTCCACCTGGCGCAACCTGAAGCAAGCACCACAGTTCA 673
QY 601 tgaccagaactaccagagactcccaacctccaggtccagagaaaggccagcagc 660
DB 674 TGACCCAGAACTACCAAGACTCAACCAACCTCCAGGCTCCCAAGGAAGGCCAGCGAGC 733
QY 661 ccaagcacaacaaacaggcggagatagctgcctgctagatagcgggtttgcccacccgg 720
DB 734 CCAAGCACAACAAACAGCGGAGATAGCTGCCTGCTAGATAGCGGCTTTGCCATCCGGG 793
QY 721 catgtggccacactgccacacacagcagatgtgggtatggaacccccctctggatacagaa 780
DB 794 CATGTGGCCACACTGCCACACACCGACGAGTGTGGGTATGGAACCCCTCTGTATACAGAA 853
QY 781 ccccttctttccaaataaaaaaaatcatcca 814
DB 854 CCCCTTCTTTTCCAAATATAAAAAAATCATCCA 887
```

RESULT 12
US-09-208-718-4
Sequence 4, Application US/09208718
GENERAL INFORMATION:
APPLICANT: Lal, Preeti

APPLICANT: Hillman, Jennifer
APPLICANT: Corley, Neil
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/208,718
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/958,820
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0379 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-853-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGTUT12
CLONE: 3126479
US-09-208-718-4

Query Match 97.6%; Score 806; DB 43; Length 903;
Best Local Similarity 99.4%; Pred. No. 3e-165;
Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 aaccagaggtgccatgggttgacaatgagctggtcacagcagcactgttactgggtc 60
DB 74 AACACAGAGGTGCCATGGGTGGACATGAGGCTGTGTACAGCAGCAGCTGTTACTGGGTC 133
QY 61 tcatgatgtgttcaactgagcagagagatgagaacagccgtgtgcccattgagccctcc 120
DB 134 TCATGATGTGGTCACTGGAGAGAGAGATGAGAACAGCCGCTGTGCCATGAGGCCCTCT 193
QY 121 tggacagagacacccctctttgcccagggccttgaagtcttaccagagagttggggaaca 180
DB 194 TGGACAGGAGACACCCCTCTTTTGGCAGGGCCTTCAAGTTTTCTACCCAGAGATTGGGGAACA 253
QY 181 ttggttgcgaaggtgttctctgattgtaaacactacagacagaagatcacctcctgagtg 240
DB 254 TTGGCTGCAAGGTGTTCCTGATTGTAACAACTACAGACAGAAAGATCACCTCCTGGATGG 313
QY 241 agccgatagtcgaagttcccggtgggcccgtgtgacgcgcgaacctatactcctggtgagtg 300
DB 314 AGCCGATAGTCAAGTTCCTGGGGGCGCTGAGCGGCGCAACCTATATCTTGGTATGGTGG 373
QY 301 atccagatgccccctagcagagcagaacccagacagagattctgagacattggtggtgtaa 360
DB 374 ATCCAGATGCCCCTAGCAGAGCAGAACCCACAGACAGAGATTCTGGAGACATTGGCTGGTAA 433
QY 361 cagatatcaagggtcccgacctgaagaaagggaagattcagggccagaggttatcagcct 420
```

Db 434 CAGATATCAAGGGCGGCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCT 493
Qy 421 accaggctccctccaccggccacacagtggtctccatcgctaccagttctttgtctatc 480
Db 494 ACCAGGCTCCCTCCCGACCGGCACACAGTGGCTTCATCGCTACCACTTCTTGTCTATC 553
Qy 481 ttcagggaaggaaggtcatctctctctctcccaagggaagaaacaaactcgaggctcttggga 540
Db 554 TTCAGGAAGGAAGATCATCTCTCTCTCCCAAGGAAGAAACAACTCGAGGCTCTTGGGA 613
Qy 541 aaatggacagatttctgaacggtttccactctccactgggggaacactgaagcaccaccagtcca 600
Db 614 AAATGGACAGATTCTGAACCGCTTCACCTGGCGCAACCTGAAGCAAGCACCCAGTTCA 673
Qy 601 tgaccagaactaccaggaactcaccacacctccaggctccagagaaagggcgagcgagc 660
Db 674 TGACCCAGAACTACCAGAGTACCACACCTCCAGGCTCCAGAGGAAGGGCCAGCGAGC 733
Qy 661 ccaagcacaacaaacaggcgagagatagctgctgctagatagccggctttggccatccggg 720
Db 734 CCAAGCACAAAACAGCGGAGATAGCTGCTGTAGATAGCGGGCTTTGCCATCCGGG 793
Qy 721 catgtgcccactgcccaccacagcagatggtggtatggaacccccctctggatacagaa 780
Db 794 CATGTGCCACACTGCCACACCGACGAGATGTGGGTATGGAACCCCTCTGGATACAGAA 853
Qy 781 ccccttcttccaataaaaaaaatcatcca 814
Db 854 CCCCTTCTTCCAAATAAAAAAATCATCCA 887

RESULT 13
US-09-390-126-4
; Sequence 4, Application US/09390126
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/390,126
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/958,820
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0379 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; IMMEDIATE SOURCE:
; LIBRARY: LUNGTUR12
; CLONE: 3126479
US-09-390-126-4

Query Match 97.6%; Score 806; DB 53; Length 903;
Best Local Similarity 99.4%; Pred. No. 3e-165;
Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 aaccagagggtgcccattgggttggaacaatgaggtggtgcacagcagcactgttactgggtc 60
Db 74 AACACAGAGGTGCCCATGGTTGGACAATAGGCTGGTCACACAGCAGCAGCTGTACTGGGTC 133
Qy 61 tcattggttggttcactggagacagagataagaacacgcccgtgtgccatagagccctcc 120
Db 134 TCATGATGGTGGTGCCTGGAGACGAGGATGAGAAACGCCCGTGTGCCATGAGGCCCTCT 193
Qy 121 tggacgaggacacacctcttttggcagggtccttgaagttttctaccagagtgggggaaca 180
Db 194 TGGACGAGGACACCTCTTTTGGCCAGGGCCTTGAAGTTTCTACCCAGAGTTGGGAACA 253
Qy 181 ttgctgcaagggttgcttctgattgtaaacactacagacagagaatcaacctcctcgtgag 240
Db 254 TTGCTGCAAGGTGTCTCTGATTGTAACAACTACAGACAGAAATCACCTCTCTGGATGG 313
Qy 241 agccatagttcaagttcccggtggggggcggtggagcggcgaacctatctcgtggtggtg 300
Db 314 AGCCGATAGTCAAGTTCCCGGGGGCGGTGGACGGCGGAACCTATATCTGTGTGATGGTGG 373
Qy 301 atccagatgcccctagcagacagaaacccacagacagagattctggagacattgggtggtaa 360
Db 374 ATCCAGATGCCCTAGCAGACAGACAGACAGAGATTCTGGAGACATTTGGCTGGTAA 433
Qy 361 cagatatacgaagggtggtgacacctgaagaaagggaagattcagggtccagaggttatacag 420
Db 434 CAGATATCAAGGGCGCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCT 493
Qy 421 aceaggtccctccaccggcgacacagtggttccatcgctaccagttcttctctatc 480
Db 494 ACCAGGCTCCCTCCCGACCGGCACACAGTGGCTTCCATCGCTACAGTTCTTGTCTATC 553
Qy 481 ttcagggaaggaaggtctatctctctcccaagggaacacaaactcgaggtccttggga 540
Db 554 TTCAGGAAGGAAGTATCTCTCTCTCTCCCAAGGAAGAAACAACTCGAGGCTTTGGGA 613
Qy 541 aaatggacagatttctgaacggtttccacctggcgaaacctgaagcagcagccagttca 600
Db 614 AAATGGACAGATTCTGAACCGCTTCCACCTGGCGAACCCTGAAGCAAGCACCCAGTTCA 673
Qy 601 tgaccagaactaccaggaactcaccacacctccaggctccagagaaagggcgagcgagc 660
Db 674 TGACCCAGAACTACCAGAGTACCACACCTCCAGGCTCCAGAGGAAGGGCCAGCGAGC 733
Qy 661 ccaagcacaacaaacaggcgagagatagctgctgctagatagcggctttgccatccggg 720
Db 734 CCAAGCACAAAACAGCGGAGATAGCTGCTGTAGATAGCGGGCTTTCGCAATCCGGG 793
Qy 721 catgtgcccactgcccaccacagcagatggtggtatggaacccccctctggatacagaa 780
Db 794 CATGTGCCACACTGCCACACCGACGAGTGTGGGTATGGAACCCCTCTGTGATACAGAA 853
Qy 781 ccccttcttccaataaaaaaaatcatcca 814
Db 854 CCCCTTCTTCCAAATAAAAAAATCATCCA 887

RESULT 14
PCT-US99-29950-61
; Sequence 61, Application PC/TUS9929950
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 47 Human Secreted Proteins

FILE REFERENCE: P2035.PCT
CURRENT APPLICATION NUMBER: PCT/US99/29950
CURRENT FILING DATE: 1999-12-16
EARLIER APPLICATION NUMBER: 60/112,809
EARLIER FILING DATE: 1998-12-17
EARLIER APPLICATION NUMBER: 60/113,006
EARLIER FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 230
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 61
LENGTH: 952
TYPE: DNA
ORGANISM: Homo sapiens
PCT-US99-29950-61

Query Match 96.4%; Score 796; DB 1; Length 952;
Best Local Similarity 99.2%; Pred. No. 4.4e-163;
Matches 821; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
QY 1 aaccagaggtgccccatgggttggaacatgaggtggttcacagcagcactgttactgggtc 60
DB 49 aaccagaggtgccccatgggttggaacatgaggtggttcacagcagcactgttactgggtc 108
QY 61 tcattgatggtggtcactggagacagagatgagaaacagcccggtgccccatgagccctcc 120
DB 109 tcattgatggtggtcactggagacagagatgagaaacagcccggtgccccatgagccctcc 168
QY 121 tggacagaggacaccccttttgcagagcccttgaagtgttctccacagagtgagggaaca 180
DB 169 tggacagaggacaccccttttgcagagcccttgaagtgttctccacagagtgagggaaca 228
QY 181 ttggctcagaggtgttcctgattgttaacactacagacagaaagatcacctctggtggtg 240
DB 229 ttggctcagaggtgttcctgattgttaacactacagacagaaagatcacctctggtggtg 288
QY 241 -agccgatagtcgaattcccgccggcggtgacggcgcaacctat-atcctggtgatggt 298
DB 289 agccgatagtcgaattcccgccggcggtgacggcgcaacctatatactggtggtggt 348
QY 299 ggtaccagatgccccctagcagagacagaaacccagacagagattctgagacattggctggt 358
DB 349 ggtaccagatgccccctagcagagacagaaacccagacagagattctgagacattggctggt 408
QY 359 aacagatatcaaggcgccagacctgaagaagaaagagattcagggccagagttatcagc 418
DB 409 aacagatatcaaggcgccagacctgaagaagaaagagattcagggccagagttatcagc 468
QY 419 ctaccaggtccctcccccagccgacacagtggtctccatgctacagattcttctgcta 478
DB 469 ctaccaggtccctcccccagccgacacagtggtctccatgctacagattcttctgcta 528
QY 479 tcttcaggaaggaaggttcattctctctctccaaaggaaacaaactcgaggtcttg 538
DB 529 tcttcaggaaggaaggttcattctctctctccaaaggaaacaaactcgaggtcttg 588
QY 539 gaaatggacagattctgaacgttttccactggtggaacctgaagcaagcaccagtt 598
DB 589 gaaatggacagattctgaacgttttccactggtggaacctgaagcaagcaccagtt 648
QY 599 catgacccgaactacaggaactcaccacacctccaggtctccagagaaagggccagcga 658
DB 649 catgacccgaactacaggaactcaccacacctccaggtctccagagaaagggccagcga 708
QY 659 gcccaagcaaaaaacagcgagatagctgctctagatagcggcttggccatccg 718
DB 709 gcccaagcaaaaaacagcgagatagctgctctagatagcggcttggccatccg 768
QY 719 ggcattgtgccacactgcccacccagcagatgtgggtatggaaacccctctggatacag 778
DB 769 ggcattgtggccacactgcccacccagcagatgtgggtatggaaacccctctggatacag 828
QY 779 aacccttcttcccaataaaaaaaatcatccaaaaaa 826

DB 829 aacccttcttcccaataaaaaaaatcatccaaaaaa 876
RESULT 15
US-09-491-404-1525
Sequence 1525, Application US/09491404
GENERAL INFORMATION:
APPLICANT: Tang, Yuanhua T.
APPLICANT: Tillinghast, John
APPLICANT: Sinku, Ankura
APPLICANT: Liu, Chenghua
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Novel Contigs Obtained
FILE REFERENCE: 785
CURRENT APPLICATION NUMBER: US/09/491,404
CURRENT FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 3796
SOFTWARE: pt_sp_genes Version 1.0
SEQ ID NO 1525
LENGTH: 886
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: sig_peptide
LOCATION: (70)...(205)
OTHER INFORMATION: this location contains the signal peptide sequence,
OTHER INFORMATION: MGMTMLRLVTAALLGLMMVVTGDESDENSPCAHEALLDFTLFCQG, Run with Signal
FEATURE:
NAME/KEY: misc_feature
LOCATION: (304)...(700)
OTHER INFORMATION: similar to g1510339 in the genepept database release 114,
OTHER INFORMATION: Run with FASTXY 3.3c00, default parameters
US-09-491-404-1525

Query Match 92.4%; Score 763; DB 56; Length 886;
Best Local Similarity 98.8%; Pred. No. 6.3e-156;
Matches 821; Conservative 0; Mismatches 5; Indels 5; Gaps 5;
QY 1 aaccagaggtgccccatgggttggaacatgaggtggttcacagcagcactgttactgggtc 60
DB 56 aaccagaggtgccccatgggttggaacatgaggtggttcacagcagcactgttactgggtc 115
QY 61 tcattgatggtggtcactggagacagagatgagaaacagcccggtgccccatgagccctcc 120
DB 116 tcattgatggtggtcactggagacagagatgagaaacagcccggtgccccatgagccctcc 175
QY 121 tggacagagacacccctcttttgcagggcgcttgaagtgttctaccagagttggggaaca 180
DB 176 tggacagagacacccctcttttgcagggcgcttgaagtgttctaccagagttggggaaca 235
QY 181 ttggctcagaggtgttctctgattgttaacactacagacagatcacctctggatgg 240
DB 236 ttggctcagaggtgttctctgattgttaacactacagacagatcacctctggatgg 295
QY 241 agccgatagtcgaattcccgccggcggtggaacgcaacctatctctggtggtggtg 300
DB 296 agccgatagtcgaattcccgccggcggtggaacgcaacctatctctggtggtggtg 355
QY 301 atccagatgccccctagcagagacagaaacccagacagagattctggagacattggcggtaa 360
DB 356 atccagatgccccctagcagagacagaaacccagacagagattctggagacattggcggtaa 415
QY 361 cagatatcaagggcgccagacctgaagaagaaagagattcagggcagagttatcag-cc 419
DB 416 cagatatcaagggcgccagacctgaagaagaaagagattcagggcagagttatcagccc 475
QY 420 taccaggtctccct-ccccacccgacacagtggt-cttccatcgctacc-agttcttgc 476
DB 476 taccaggtctccctccccacccgacacagtggtggtcttccatcgctaccagagttcttgc 535

THIS PAGE BLANK (USPTO)